Solutions 1

Prof. Krishna R. Pattipati

Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

ECE 6141
Neural Networks for Classification and Optimization

- Problem 1

$$(i)\ \sigma(x)=\frac{1}{1+e^{-x}};(1-\sigma(x))=\frac{e^{-x}}{1+e^{-x}}=\frac{1}{1+e^{x}}=\sigma(-x)$$

$$(ii)\frac{d\sigma(x)}{dx}=\sigma(x)(1-\sigma(x))=\sigma(x)\sigma(-x)$$

$$\frac{d\sigma(x)}{dx}=\frac{e^{-x}}{\left(1+e^{-x}\right)^{2}}=\sigma(x)\sigma(-x)=\sigma(x)(1-\sigma(x))$$

$$(iii)\ln\sigma(x)=-\varsigma(-x)$$

$$\varsigma(x)=\ln(1+\exp(x))\Rightarrow\varsigma(-x)=\ln(1+\mathrm{e}^{-x})=-\ln\sigma(x)$$

$$(iv)\frac{d\varsigma(x)}{dx}=\sigma(x)$$

$$\varsigma(x)=\ln(1+\exp(x))\Rightarrow\frac{d\varsigma(x)}{dx}=\frac{e^{x}}{1+e^{x}}=\frac{1}{1+e^{-x}}=\sigma(x)$$

$$(v)x=\ln\frac{\sigma(x)}{1-\sigma(x)};\sigma(x)\in(0,1);x\in(-\infty,\infty)$$

$$\frac{\sigma(x)}{1-\sigma(x)}=\frac{1}{1+e^{-x}}\frac{1+e^{-x}}{e^{-x}}=e^{x}\Rightarrow\ln\frac{\sigma(x)}{1-\sigma(x)}=x$$

$$(vi)x=\ln(\exp(\varsigma(x))-1);\varsigma(x)\in(0,\infty);x\in(-\infty,\infty)$$

$$e^{\varsigma(x)}-1=1+e^{x}-1=e^{x}\Rightarrow\ln(e^{\varsigma(x)}-1)=x$$

$$(vii)\varsigma(x)-\varsigma(-x)=-\ln\sigma(-x)+\ln\sigma(x)=\ln\frac{\sigma(x)}{\sigma(-x)}=x$$

$$\varsigma(x)-\varsigma(-x)=\ln(1+e^{x})-\ln(1+e^{-x})=-\ln\sigma(-x)+\ln\sigma(x)=\ln\frac{\sigma(x)}{\sigma(-x)}=x$$

- Problem 1

$$b)\, i.\ f(\underline{w}) = \sum_{n=1}^{N} (z_n - \underline{w}^T \underline{x}_n)^2 = \sum_{n=1}^{N} e_n^2$$

$$\nabla_{\underline{w}} f = -2\sum_{n=1}^{N} (z_n - \underline{w}^T \underline{x}_n)\underline{x}_n = -2\sum_{n=1}^{N} e_n \underline{x}_n$$

$$\nabla_{\underline{w}}^2 f = 2\sum_{n=1}^{N} \underline{x}_n \underline{x}_n^T > 0$$

$$ii.\, f(\underline{w}) = \sum_{n=1}^{N} (z_n - y_n)^2 = \sum_{n=1}^{N} e_n^2 \ \ where\ \ y_n = g(\underline{w}^T \underline{x}_n) = \frac{1}{1+e^{-\underline{w}^T \underline{x}_n}}$$

$$\nabla_{\underline{w}} f = -2\sum_{n=1}^{N} (z_n - y_n) y_n (1-y_n)\underline{x}_n = -2\sum_{n=1}^{N} e_n y_n (1-y_n)\underline{x}_n$$

$$\nabla_{\underline{w}}^2 f = 2\sum_{n=1}^{N} y_n (1-y_n)[y_n(1-y_n) - e_n(1-2y_n)]\, x_n \underline{x}_n^T$$

- Problem 1

$$b)\,iii.\; f(\underline{w}) = -\sum_{n=1}^{N}[z_n \ln y_n + (1-z_n)\ln(1-y_n)]\; where\; y_n = g(\underline{w}^T \underline{x}_n) = \frac{1}{1+e^{-\underline{w}^T \underline{x}_n}}$$

$$\nabla_{\underline{w}} f = -\sum_{n=1}^{N}[\frac{z_n}{y_n} - \frac{1-z_n}{1-y_n}]\, y_n(1-y_n)\underline{x}_n$$

$$= -\sum_{n=1}^{N}[z_n - y_n]\underline{x}_n$$

$$= -\sum_{n=1}^{N} e_n \underline{x}_n ; e_n = z_n - y_n$$

$$\nabla_{\underline{w}}^2 f = \sum_{n=1}^{N} y_n(1-y_n)\underline{x}_n \underline{x}_n^T > 0$$

$$A \subset B \Rightarrow AB = A \Rightarrow \overline{B} \subset \overline{A} \Rightarrow \overline{A}\overline{B} = \overline{B}$$
$$xy = x \Rightarrow (1-x)(1-y) = 1 - x - y + xy = 1 - y$$

- Problem 2

$$a)\ P(B = True \mid A = True) = 1 \Rightarrow P(B = false \mid A = True) = 0$$

$$P(A = False \mid B = False) = 1 - P(A = True \mid B = False)$$
$$= 1 - \frac{P(B = False \mid A = True)P(A = True)}{P(B = False)} = 1$$

$b)$

B	C	P(A=1|B,C)	P(A=0|B,C)
0	0	0.10	0.90
0	1	0.99	0.01
1	0	0.80	0.20
1	1	0.25	0.75

P(B=1)=0.65
P(C=1)=0.77

$$P(B=1 \mid A=0) = \frac{P(A=0 \mid B=1)P(B=1)}{P(A=0)} = \frac{P(A=0 \mid B=1)P(B=1)}{P(A=0 \mid B=1)P(B=1) + P(A=0 \mid B=0)P(B=0)}$$

$$P(A=0 \mid B=1) = \sum_{C=0}^{1} P(A=0, C \mid B=1) = \sum_{C=0}^{1} P(A=0 \mid B=1, C)P(C \mid B=1)$$
$$= \sum_{C=0}^{1} P(A=0 \mid B=1, C)P(C) = 0.20 * 0.23 + 0.75 * 0.77 = 0.6235$$

- Problem 2

Similarly, $P(A=0 \mid B=0) = \sum_{C=0}^{1} P(A=0 \mid B=0, C)P(C) = 0.90*0.23 + 0.01*0.77 = 0.2147$

$$P(B=1 \mid A=0) = \frac{0.6235*0.65}{0.6235*0.65 + 0.2147*0.35} = 0.8436$$

$$P(B=0 \mid A=0) = 1 - P(B=1 \mid A=0) = 0.1564$$

$$P(C=1 \mid A=0) = \frac{P(A=0 \mid C=1)P(C=1)}{P(A=0 \mid C=1)P(C=1) + P(A=0 \mid C=0)P(C=0)}$$

$$P(A=0 \mid C=1) = \sum_{B=0}^{1} P(A=0 \mid C=1, B)P(B) = 0.01*0.35 + 0.75*0.65 = 0.4910$$

$$P(A=0 \mid C=0) = \sum_{B=0}^{1} P(A=0 \mid C=0, B)P(B) = 0.90*0.35 + 0.20*0.65 = 0.4450$$

$$P(C=1 \mid A=0) = \frac{0.4910*0.77}{0.4910*0.77 + 0.4450*0.23} = 0.7870$$

$$P(C=0 \mid A=0) = 1 - P(C=1 \mid A=0) = 0.2130$$

- Problem 2

$$P(B=1|A=1)=\frac{P(A=1|B=1)P(B=1)}{P(A=1|B=1)P(B=1)+P(A=1|B=0)P(B=0)}$$

$$P(A=1|B=1)=1-P(A=0|B=1)=1-0.6235=0.3765$$

$$P(A=1|B=0)=1-P(A=0|B=0)=1-0.2147=0.7853$$

$$P(B=1|A=1)=\frac{0.3765*0.65}{0.3765*0.65+0.7853*0.35}=0.4710$$

$$P(B=0|A=1)=1-P(B=1|A=1)=0.5290$$

$$P(C=1|A=1)=\frac{P(A=1|C=1)P(C=1)}{P(A=1|C=1)P(C=1)+P(A=1|C=0)P(C=0)}$$

$$P(A=1|C=1)=1-P(A=0|C=1)=1-0.4910=0.5090$$

$$P(A=1|C=0)=1-P(A=0|C=0)=1-0.4450=0.5550$$

$$P(C=1|A=1)=\frac{0.5010*0.77}{0.5010*0.77+0.5550*0.23}=0.7543$$

$$P(C=0|A=1)=1-P(C=1|A=1)=0.2457$$

Problem 2(c)

$$Events$$

$$Bags: C_1, C_2$$

$$Black\ \ Balls: B,\ B_1, B_2$$

$$White\ Ball: W$$

$$Need: P(B_2 \mid B_1)\ and\ P(W \mid B_1)$$

$$Know\ P(C_1) = P(C_2) = 1/2;$$

$$P(B \mid C_1) = 1; P(B \mid C_2) = P(W \mid C_2) = 1/2$$

$$So, P(B_2 \mid B_1) = P(B_2, C_1 \mid B_1) + P(B_2, C_2 \mid B_1)$$

$$= P(B_2 \mid C_1, B_1) P(C_1 \mid B_1) + P(B_2 \mid C_2, B_1) P(C_2 \mid B_1)$$

$$= 1.P(C_1 \mid B_1) + 0.P(C_2 \mid B_1)$$

$$P(C_1 \mid B_1) = \frac{P(B_1 \mid C_1)P(C_1)}{P(B_1)} = \frac{P(B \mid C_1)P(C_1)}{P(B)}$$

$$= \frac{P(B \mid C_1)P(C_1)}{P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2)} = \frac{1.(1/2)}{1.(1/2) + (1/2).(1/2)} = \frac{2}{3}$$

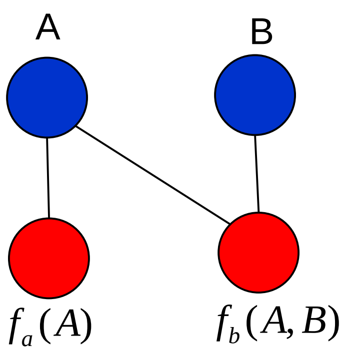
$$P(W \mid B_1) = 1 - P(B_2 \mid B_1) = \frac{1}{3}$$

Can also get this by drawing a tree of possibilities.

- Problem 2(d)

$$f_a(A) = P(A) = \begin{matrix} A=T \\ A=F \end{matrix}\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \quad P(A,B) = P(A)P(B|A) = f_a(A)\,f_b(A,B)$$

$$f_b(A,B) = P(B|A) = \begin{matrix} \\ A=T \\ A=F \end{matrix}\begin{matrix} B=T & B=F \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix}$$



Evidence, $B=F$. What is $P(A|B=F)$?

Treat A as the root.

$$\mu_{B\to f_b}(B=F)=1;\ \mu_{B\to f_b}(B=T)=0$$

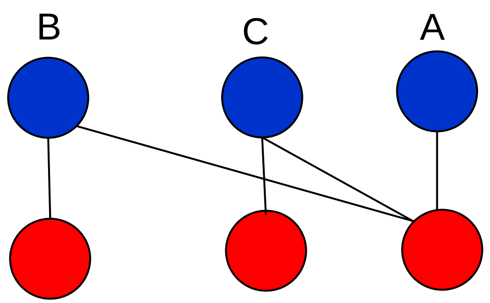
$$\mu_{f_b\to A}(A) = \sum_B f_b(A,B)\mu_{B\to f_b}(B) \Rightarrow \mu_{f_b\to A}(A=T)=0;\ \mu_{f_b\to A}(A=F)=1$$

$$\mu_{f_a\to A}(A) = P(A) \Rightarrow \mu_{f_a\to A}(A=T)=0.5;\ \mu_{f_a\to A}(A=F)=0.5$$

$$P(A|B=F) = \mu_{f_a\to A}(A)\mu_{f_b\to A}(A) \Rightarrow \begin{bmatrix} \mu_{f_a\to A}(A=T)\mu_{f_b\to A}(A=T) \\ \mu_{f_a\to A}(A=F)\mu_{f_b\to A}(A=F) \end{bmatrix} = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \Rightarrow Normalize: \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Problem 2(d)

$$f_b(B) = P(B) = \begin{matrix} B=1 \\ B=0 \end{matrix}\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix}; \; f_c(C) = P(C) = \begin{matrix} C=1 \\ C=0 \end{matrix}\begin{bmatrix} 0.77 \\ 0.23 \end{bmatrix}$$



$$f_a(A,B,C) = P(A|B,C) = \begin{matrix} & \\ (B,C)=(0,0) \\ (B,C)=(0,1) \\ (B,C)=(1,0) \\ (B,C)=(1,1) \end{matrix}\begin{matrix} A=1 & A=0 \\ \left[\begin{matrix} 0.1 & 0.9 \\ 0.99 & 0.01 \\ 0.80 & 0.20 \\ 0.25 & 0.75 \end{matrix}\right] \end{matrix}$$

$$P(A,B,C) = P(B)P(C)P(A|B,C) = f_b(B)\,f_c(C)\,f_a(A,B,C)$$

Evidence, A = 0. What is $P(B=1|A=0)$ and $P(C=1|A=0)$? Make A root.

Leaves $\rightarrow$ Root

$$\mu_{f_b\to B}(B=1) = 0.65; \mu_{f_b\to B}(B=0) = 0.35$$

$$\mu_{f_c\to C}(C=1) = 0.77; \mu_{f_c\to C}(C=0) = 0.23$$

$$\mu_{B\to f_a}(B=1) = \mu_{f_b\to B}(B=1) = 0.65; \mu_{B\to f_a}(B=0) = \mu_{f_b\to B}(B=0) = 0.35$$

$$\mu_{C\to f_a}(C=1) = \mu_{f_c\to C}(C=1) = 0.77; \mu_{C\to f_a}(C=0) = \mu_{f_c\to C}(C=0) = 0.23$$

$$\mu_{f_a\to A}(A) = \sum_{B=0}^{1}\sum_{C=0}^{1} f_a(A,B,C)\mu_{B\to f_a}(B)\mu_{C\to f_a}(C)$$

$$\mu_{f_a\to A}(A=1) = \sum_{B=0}^{1}\sum_{C=0}^{1} f_a(1,B,C)\mu_{B\to f_a}(B)\mu_{C\to f_a}(C)$$

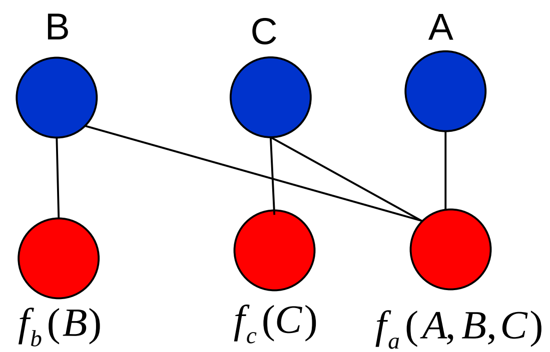
$$= (0.1)(0.35)(0.23) + (0.99)(0.35)(0.77) + (0.80)(0.65)(0.23) + (0.25)(0.65)(0.77) = 0.51958$$

$$\mu_{f_a\to A}(A=0) = \sum_{B=0}^{1}\sum_{C=0}^{1} f_a(0,B,C)\mu_{B\to f_a}(B)\mu_{C\to f_a}(C)$$

$$= (0.9)(0.35)(0.23) + (0.01)(0.35)(0.77) + (0.20)(0.65)(0.23) + (0.75)(0.65)(0.77) = 0.48042$$

Problem 2(d)

$$f_b(B) = \begin{matrix} B=1 \\ B=0 \end{matrix}\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix}; \; f_c(C) = \begin{matrix} C=1 \\ C=0 \end{matrix}\begin{bmatrix} 0.77 \\ 0.23 \end{bmatrix}$$



$$f_a(A,B,C) = P(A\,|\,B,C) = \begin{matrix} \\ (B,C)=(0,0) \\ (B,C)=(0,1) \\ (B,C)=(1,0) \\ (B,C)=(1,1) \end{matrix}\begin{matrix} A=1 & A=0 \\ 0.1 & 0.9 \\ 0.99 & 0.01 \\ 0.80 & 0.20 \\ 0.25 & 0.75 \end{matrix}$$

Evidence, A = 0. What is $P(B=1\,|\,A=0)$ and $P(C=1\,|\,A=0)$?

Root $\to$ Leaves; Evidence A=0

$\mu_{A\to f_a}(A=1)=0;\ \mu_{A\to f_a}(A=0)=1$

$$\mu_{f_a\to B}(B=1) = \sum_{A=0}^{1}\sum_{C=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{C\to f_a}(C) = \sum_{C=0}^{1} f_a(0,1,C)\mu_{C\to f_a}(C)$$

$= (0.2)(0.23) + (0.75)(0.77) = 0.6235$

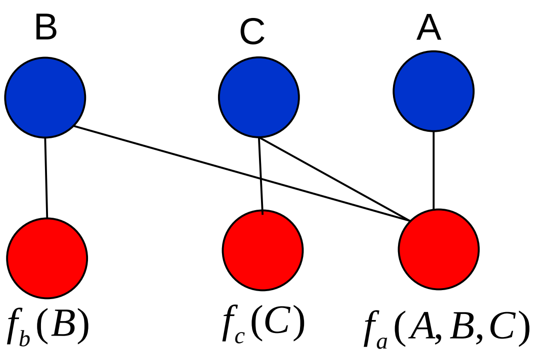
$$\mu_{f_a\to B}(B=0) = \sum_{A=0}^{1}\sum_{C=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{C\to f_a}(C) = \sum_{C=0}^{1} f_a(0,0,C)\mu_{C\to f_a}(C)$$

$= (0.9)(0.23) + (0.01)(0.77) = 0.2147$

$$P(B\,|\,A=0) = \begin{bmatrix} \mu_{f_a\to B}(B=1)\mu_{f_b\to B}(B=1) \\ \mu_{f_a\to B}(B=0)\mu_{f_b\to B}(B=0) \end{bmatrix} = \begin{bmatrix} 0.405275 \\ 0.075145 \end{bmatrix} \xrightarrow{Normalize} \begin{bmatrix} 0.8436 \\ 0.1564 \end{bmatrix}$$

- Problem 2(d)

$$f_b(B) = \begin{matrix} B=1 \\ B=0 \end{matrix}\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix}; \; f_c(C) = \begin{matrix} C=1 \\ C=0 \end{matrix}\begin{bmatrix} 0.77 \\ 0.23 \end{bmatrix}$$



$$f_a(A,B,C) = P(A \mid B,C) = \begin{matrix} \\ (B,C)=(0,0) \\ (B,C)=(0,1) \\ (B,C)=(1,0) \\ (B,C)=(1,1) \end{matrix}\begin{bmatrix} A{=}1 & A{=}0 \\ 0.1 & 0.9 \\ 0.99 & 0.01 \\ 0.80 & 0.20 \\ 0.25 & 0.75 \end{bmatrix}$$

Evidence, A = 0. What is $P(B=1 \mid A=0)$ and $P(C=1 \mid A=0)$?

Root $\rightarrow$ Leaves; Evidence A=0

$\mu_{A\to f_a}(A=1)=0; \mu_{A\to f_a}(A=0)=1$

$$\mu_{f_a\to C}(C=1) = \sum_{A=0}^{1}\sum_{B=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{B\to f_a}(B) = \sum_{B=0}^{1} f_a(0,B,1)\mu_{B\to f_a}(B)$$

$= (0.01)(0.35) + (0.75)(0.65) = 0.491$

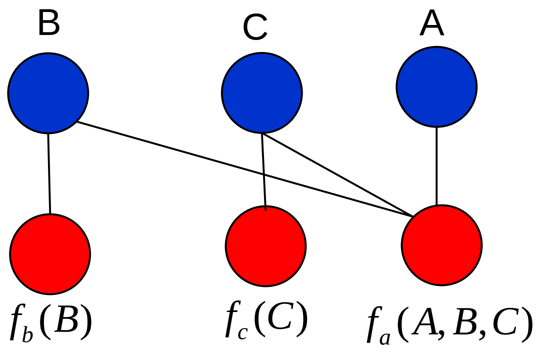
$$\mu_{f_a\to C}(C=0) = \sum_{A=0}^{1}\sum_{B=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{B\to f_a}(B) = \sum_{B=0}^{1} f_a(0,B,0)\mu_{B\to f_a}(B)$$

$= (0.90)(0.35) + (0.20)(0.65) = 0.445$

$$P(C \mid A=0) = \begin{bmatrix} \mu_{f_a\to C}(C=1)\mu_{f_c\to C}(C=1) \\ \mu_{f_a\to C}(C=0)\mu_{f_c\to C}(C=0) \end{bmatrix} = \begin{bmatrix} 0.37807 \\ 0.10235 \end{bmatrix} \xrightarrow{Normalize} \begin{bmatrix} 0.787 \\ 0.213 \end{bmatrix}$$

Problem 2(d)

$$f_b(B) = \begin{matrix} B=1 \\ B=0 \end{matrix}\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix};\ f_c(C) = \begin{matrix} C=1 \\ C=0 \end{matrix}\begin{bmatrix} 0.77 \\ 0.23 \end{bmatrix}$$



$$f_a(A,B,C) = P(A\,|\,B,C) = \begin{matrix} \\ (B,C)=(0,0) \\ (B,C)=(0,1) \\ (B,C)=(1,0) \\ (B,C)=(1,1) \end{matrix}\begin{matrix} A{=}1 & A{=}0 \\ 0.1 & 0.9 \\ 0.99 & 0.01 \\ 0.80 & 0.20 \\ 0.25 & 0.75 \end{matrix}$$

Evidence, A = 0. What is $P(B=1\,|\,A=0)$ and $P(C=1\,|\,A=0)$?

Root $\rightarrow$ Leaves; Evidence A=1

$\mu_{A\to f_a}(A=1)=1;\ \mu_{A\to f_a}(A=0)=0$

$$\mu_{f_a\to B}(B=1) = \sum_{A=0}^{1}\sum_{C=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{C\to f_a}(C) = \sum_{C=0}^{1} f_a(1,1,C)\mu_{C\to f_a}(C)$$

$$=(0.8)(0.23)+(0.25)(0.77)=0.3765$$

$$\mu_{f_a\to B}(B=0) = \sum_{A=0}^{1}\sum_{C=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{C\to f_a}(C) = \sum_{C=0}^{1} f_a(1,0,C)\mu_{C\to f_a}(C)$$

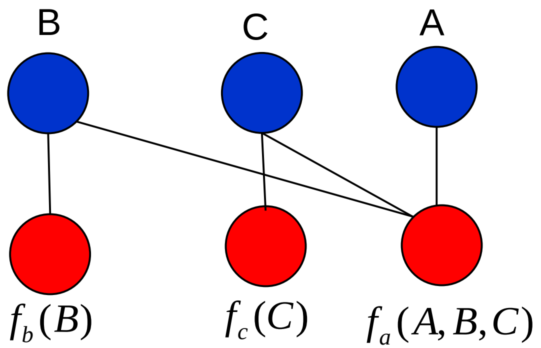
$$=(0.1)(0.23)+(0.99)(0.77)=0.7853$$

$$P(B\,|\,A=1) = \begin{bmatrix} \mu_{f_a\to B}(B=1)\mu_{f_b\to B}(B=1) \\ \mu_{f_a\to B}(B=0)\mu_{f_b\to B}(B=0) \end{bmatrix} = \begin{bmatrix} 0.2447 \\ 0.2749 \end{bmatrix} \xrightarrow{Normalize} \begin{bmatrix} 0.4710 \\ 0.5290 \end{bmatrix}$$

Problem 2(d)

$$f_b(B) = \begin{matrix} B=1 \\ B=0 \end{matrix}\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix}; \; f_c(C) = \begin{matrix} C=1 \\ C=0 \end{matrix}\begin{bmatrix} 0.77 \\ 0.23 \end{bmatrix}$$

$$f_a(A,B,C) = P(A\,|\,B,C) = \begin{matrix} \\ (B,C)=(0,0) \\ (B,C)=(0,1) \\ (B,C)=(1,0) \\ (B,C)=(1,1) \end{matrix}\begin{matrix} A{=}1 & A{=}0 \\ \left[\begin{matrix} 0.1 & 0.9 \\ 0.99 & 0.01 \\ 0.80 & 0.20 \\ 0.25 & 0.75 \end{matrix}\right] \end{matrix}$$



Evidence, A = 0. What is $P(B=1\,|\,A=0)$ and $P(C=1\,|\,A=0)$?

Root $\rightarrow$ Leaves; Evidence A=1

$\mu_{A\to f_a}(A=1)=1; \mu_{A\to f_a}(A=0)=0$

$$\mu_{f_a\to C}(C=1) = \sum_{A=0}^{1}\sum_{B=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{B\to f_a}(B) = \sum_{B=0}^{1} f_a(1,B,1)\mu_{B\to f_a}(B)$$

$$=(0.99)(0.35)+(0.25)(0.65)=0.509$$

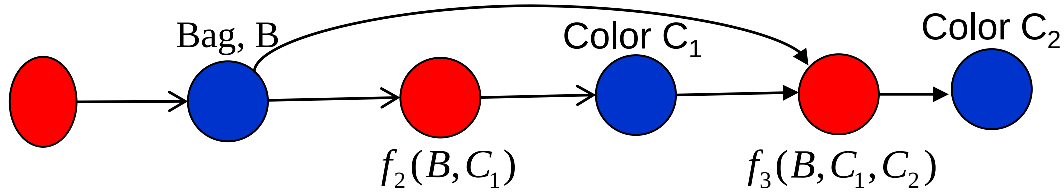
$$\mu_{f_a\to C}(C=0) = \sum_{A=0}^{1}\sum_{B=0}^{1} f_a(A,B,C)\mu_{A\to f_a}(A)\mu_{B\to f_a}(B) = \sum_{B=0}^{1} f_a(1,B,0)\mu_{B\to f_a}(B)$$

$$=(0.10)(0.35)+(0.80)(0.65)=0.555$$

$$P(C\,|\,A=1) = \begin{bmatrix} \mu_{f_a\to C}(C=1)\mu_{f_c\to C}(C=1) \\ \mu_{f_a\to C}(C=0)\mu_{f_c\to C}(C=0) \end{bmatrix} = \begin{bmatrix} 0.39193 \\ 0.12765 \end{bmatrix} \xrightarrow{Normalize} \begin{bmatrix} 0.7543 \\ 0.2457 \end{bmatrix}$$

Problem 2(c)

$$f_1(B) = P(B) = \begin{matrix} B=B_1 \\ B=B_2 \end{matrix}\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$



$$P(B,C_1,C_2) = P(B)P(C_1\,|\,B)P(C_2\,|\,B,C_1) = f_1(B)\,f_2(B,C_1)\,f_3(B,C_1,C_2)$$

$$f_2(B,C_1) = P(C_1\,|\,B) = \begin{matrix} \\ B=1 \\ B=2 \end{matrix}\begin{matrix} C_1=Bl & C_1=W \\ \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{matrix};\; f_3(B,C_1,C_2) = \begin{matrix} \\ 1Bl \\ 2Bl \\ 2W \end{matrix}\begin{matrix} C_2=Bl & C_2=W \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \end{matrix}$$

$\text{Iteration } 0: C_1 = Bl, Root = C_2$

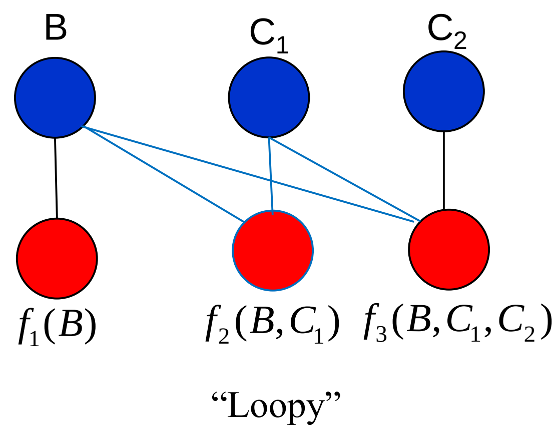
$$\mu_{f_1\to B}(B) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix};\mu_{B\to f_2}(B) = \mu_{f_1\to B}(B)\mu_{f_3\to B}(B) = \mu_{f_1\to B}(B) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$\mu_{B\to f_3}(B) = \mu_{f_1\to B}(B)\mu_{f_2\to B}(B) = \mu_{f_1\to B}(B) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$\mu_{f_2\to C_1}(C_1) = \sum_{B=1}^{2} f_2(B,C_1)\mu_{B\to f_2}(B) = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix};_1 \mu_{C_1\to f_3}(C_1) = \mu_{f_2\to C_1}(C_1) \overset{C_1=Bl}{=} \begin{bmatrix} \frac{3}{4} \\ 0 \end{bmatrix};C_1 = Bl$$

$$\mu_{f_3\to C_2}(C_2) = \sum_{B=1}^{2} f_3(B,Bl,C_2)\mu_{C_1\to f_3}(Bl)\mu_{B\to f_3}(B) = \begin{bmatrix} 1.\frac{3}{4}.\frac{1}{2}+1.\frac{3}{4}.\frac{1}{2} \\ 1.\frac{3}{4}.\frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{3}{8} \end{bmatrix}$$

Iterations will stabilize. $P(C_2\,|\,C_1 = Bl) = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$



- Problem 3

$$Ga(x;a,b)=\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}; x>0; a>0; b>0$$

- $E(x)=\frac{b^a}{\Gamma(a)}\int_0^\infty x^a e^{-bx}dx=\frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1)}{b^{a+1}}=\frac{a}{b}$
- $E(x^2)=\frac{b^a}{\Gamma(a)}\int_0^\infty x^{a+1} e^{-bx}dx=\frac{b^a}{\Gamma(a)}\frac{\Gamma(a+2)}{b^{a+2}}=\frac{a(a+1)}{b^2}$
- $Var(x)=E(x^2)-[E(x)]^2=\frac{a(a+1)}{b^2}-\frac{a^2}{b^2}=\frac{a}{b^2}$
- $\text{mode}(x)\Rightarrow Ga(x;a,b)\cdots is\ \max imum$

$$\frac{\partial Ga(x;a,b)}{\partial x}=\frac{b^a}{\Gamma(a)}[(a-1)x^{a-2}e^{-bx}-bx^{a-1}e^{-bx}]=0$$

$$\Rightarrow (a-1)-bx=0\Rightarrow \text{mode}(x)=\frac{a-1}{b}\ \ for\ a>1$$

- Problem 3

$$Ga(x;a,b)=\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx};\, x>0;\, a>0;\, b>0$$

- $Entropy\ H(x)=-\int_0^\infty Ga(x;a,b)\ln[Ga(x;a,b)]dx$

$$=-\frac{b^a}{\Gamma(a)}\int_0^\infty x^{a-1}e^{-bx}\left(a\ln b-\ln\Gamma(a)+(a-1)\ln x-bx\right)dx$$

$$=-a\ln b+\ln\Gamma(a)-(a-1)E[\ln x]+a$$

$$To\ get\ E[\ln x],\ note\ G(x;a,b)=e^{a\ln b-\ln\Gamma(a)+(a-1)\ln x-bx}=e^{\begin{bmatrix}(a-1) & -b\end{bmatrix}\begin{bmatrix}\ln x\\ x\end{bmatrix}-\underbrace{(\ln\Gamma(a)-a\ln b)}_{A(a,b)}}$$

$$E(\ln x)=\frac{\partial A(a,b)}{\partial a}=\frac{d\ln\Gamma(a)}{da}-\ln b=\Psi(a)-\ln b;\ \Psi(a)=\frac{d\ln\Gamma(a)}{da}\text{.... } digamma\ function$$

$$Aside: E(x)=-\frac{\partial A(a,b)}{\partial b}=\frac{a}{b};\mathrm{cov}(\ln x,x)=\begin{bmatrix}\frac{d^2\ln\Gamma(a)}{da^2} & \frac{1}{b}\\ \frac{1}{b} & \frac{a}{b^2}\end{bmatrix}$$

$$so,\ H(x)=-a\ln b+\ln\Gamma(a)-(a-1)[\Psi(a)-\ln b]+a=\ln\Gamma(a)-(a-1)\Psi(a)-\ln b+a$$

$$=a+\ln\left(\frac{\Gamma(a)}{b}\right)-(a-1)\Psi(a)$$

Problem 3

2) $Y = \dfrac{1}{X}$

$$Approach\,1: P_y(Y \le y) = P_x(\frac{1}{X} \le y) = P_x(X \ge \frac{1}{y}) = \frac{b^a}{\Gamma(a)} \int_{1/y}^{\infty} x^{a-1} e^{-bx} dx$$

$$f_y(y) = \frac{dP_y(Y \le y)}{dy} = \frac{b^a}{\Gamma(a)} \frac{1}{y^2} \left(\frac{1}{y}\right)^{a-1} e^{-b/y} = \frac{b^a}{\Gamma(a)} (y)^{-a-1} e^{-b/y} = IG(y; a, b)$$

Leibniz's rule for differentiating integrals

$$Approach\,2: f_y(y) = f_x(x)|_{x=1/y} |\frac{dx}{dy}| = \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{a-1} e^{-b/y} \frac{1}{y^2} = \frac{b^a}{\Gamma(a)} (y)^{-a-1} e^{-b/y} = IG(y; a, b)$$

- $Entropy\ H(y) = -\int_0^{\infty} IG(y; a, b) \ln[IG(y; a, b)] dy$

$$= -\frac{b^a}{\Gamma(a)} \int_0^{\infty} y^{-a-1} e^{-b/y} \left(a \ln b - \ln\Gamma(a) - (a+1) \ln y - \frac{b}{y} \right) dy$$

$$= \text{-}\, a \ln b + \ln\Gamma(a) + (a+1) E[\ln y] + bE\left(\frac{1}{y}\right)$$

$$To\ get\ E[\ln y],\ note\ IG(y; a, b) = e^{a \ln b - \ln\Gamma(a) - (a+1)\ln y - b/y} = e^{\begin{bmatrix} -(a+1) & -b \end{bmatrix} \begin{bmatrix} \ln y \\ 1/y \end{bmatrix} - \underbrace{(\ln\Gamma(a) - a \ln b)}_{A(a,b)}}$$

$$E(\ln y) = -\frac{\partial A(a,b)}{\partial a} = -\frac{d \ln\Gamma(a)}{da} + \ln b = -\Psi(a) + \ln b;\ \ \Psi(a) = \frac{d \ln\Gamma(a)}{da} \ldots.\ digamma\ function$$

$$E(1/y) = -\frac{\partial A(a,b)}{\partial b} = \frac{a}{b}$$

$$so,\ H(x) = -a \ln b + \ln\Gamma(a) - (a+1)[\Psi(a) - \ln b] + a = \ln\Gamma(a) - (a+1)\Psi(a) + \ln b + a$$

$$= a + \ln[b\Gamma(a)] - (a+1)\Psi(a)$$

- Problem 4

$$Beta(\theta;a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1};\theta\in[0,1];a>0;b>0$$

$$\bullet\ E(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta^{a}(1-\theta)^{b-1}d\theta=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}=\frac{a}{a+b}$$

$$\bullet\ E(\theta^2)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta^{a+1}(1-\theta)^{b-1}d\theta=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)}=\frac{a(a+1)}{(a+b)(a+b+1)}$$

$$\bullet\ Var(\theta)=E(\theta^2)-[E(\theta)]^2=\frac{a(a+1)}{(a+b)(a+b+1)}-\frac{a^2}{(a+b)^2}$$

$$=\frac{a}{(a+b)}\left(\frac{a+1}{a+b+1}-\frac{a}{a+b}\right)=\frac{ab}{(a+b)^2(a+b+1)}$$

$$\bullet\ \mathrm{mod}\,e(x)\Rightarrow Beta(\theta;a,b)\cdots is\ \max imum$$

$$\frac{\partial Beta(\theta;a,b)}{\partial\theta}=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}[(a-1)\theta^{a-2}(1-\theta)^{b-1}-(b-1)\theta^{a-1}(1-\theta)^{b-2}]=0$$

$$\Rightarrow(a-1)(1-\theta)-(b-1)\theta=0\Rightarrow\mathrm{mod}\,e(\theta)=\frac{a-1}{a+b-2}\ \ for\ a>1;b>1$$

- Problem 4

$$Beta(\theta;a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1};\theta\in[0,1];a>0;b>0$$

- $$H(\theta)=-\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta^{a}(1-\theta)^{b-1}\left[\ln\Gamma(a+b)-\ln\Gamma(a)-\ln\Gamma(b)+(a-1)\ln\theta+(b-1)\ln(1-\theta)\right]d\theta$$

$$=\ln\Gamma(a)+\ln\Gamma(b)-\ln\Gamma(a+b)-(a-1)E(\ln\theta)-(b-1)E\{\ln(1-\theta)\}$$

$$Note: Beta(\theta;a,b)=e^{\left[(a-1)\quad(b-1)\right]\begin{bmatrix}\ln\theta\\ \ln(1-\theta)\end{bmatrix}-\underbrace{[\ln\Gamma(a)+\ln\Gamma(b)-\ln\Gamma(a+b)]}_{A(a,b)}}$$

$$E(\ln\theta)=\frac{\partial A(a,b)}{\partial a}=\Psi(a)-\Psi(a+b)$$

$$E(\ln(1-\theta))=\frac{\partial A(a,b)}{\partial b}=\Psi(b)-\Psi(a+b)$$

$$so,\ H(\theta)=\ln\Gamma(a)+\ln\Gamma(b)-\ln\Gamma(a+b)-(a-1)\Psi(a)-(b-1)\Psi(b)+(a+b-2)\Psi(a+b)$$

$$Aside: \mathrm{cov}[\ln\theta,\ln(1-\theta)]=\begin{bmatrix}\frac{\partial\Psi(a)}{\partial a}-\frac{\partial\Psi(a+b)}{\partial a} & -\frac{\partial\Psi(a+b)}{\partial b}\\ -\frac{\partial\Psi(a+b)}{\partial b} & \frac{\partial\Psi(b)}{\partial b}-\frac{\partial\Psi(a+b)}{\partial b}\end{bmatrix};\frac{\partial\Psi(a+b)}{\partial b}=\frac{\partial\Psi(a+b)}{\partial a}$$

The same process applies to Dirichlet ditribution

- Problem 5

$$KL(p\|q) = -\int p(x)\ln\{\frac{q(x)}{p(x)}\}dx = -E_p\{\ln q(x)\} - H_p(x)$$

$$p(x) = N(\mu,\sigma^2); q(x) = N(m,s^2)$$

$$\ln p(x) = -\frac{1}{2}\ln(2\pi) - \ln\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\ln q(x) = -\frac{1}{2}\ln(2\pi) - \ln s - \frac{(x-m)^2}{2s^2}$$

$$KL(q\|p) = -E_q\{\ln p(x)\} - H_q(x)$$
$$\neq KL(p\|q)$$
$\Rightarrow$ KL is not symmetric

$$\Rightarrow KL(p\|q) = -E_p\{\ln\frac{\sigma}{s} + \frac{(x-\mu)^2}{2\sigma^2} - \frac{(x-m)^2}{2s^2}\} = \ln\frac{s}{\sigma} - \frac{1}{2} + \frac{\mu^2+\sigma^2-2\mu m+m^2}{2s^2}$$

$$= \ln\frac{s}{\sigma} - \frac{1}{2} + \frac{(\mu-m)^2+\sigma^2}{2s^2} = \frac{1}{2}[\ln\frac{s^2}{\sigma^2} + \frac{(\mu-m)^2}{s^2} + \frac{\sigma^2}{s^2} - 1]$$

Multivariate case (n vectors): $\frac{1}{2}[\ln\frac{\det(S)}{\det(\Sigma)} + (\underline{\mu}-\underline{m})^T S^{-1}(\underline{\mu}-\underline{m}) + tr(S^{-1}\Sigma) - n]$

Minimum when $\underline{\mu} = \underline{m}$ and $S = \Sigma \Rightarrow$ moment matching

$$KL(p\|q) = \frac{1}{2}[\ln\frac{\det(S)}{\det(\Sigma)} + (\underline{\mu}-\underline{m})^T S^{-1}(\underline{\mu}-\underline{m}) + tr(S^{-1}\Sigma) - n] \text{ for Gaussian}$$

Suppose $p(\underline{x})$ is sum of Gaussian $\Rightarrow p(\underline{x}) = \sum_{i=1}^{L}\alpha_i N(\underline{\mu}_i, \Sigma_i); \sum_{i=1}^{L}\alpha_i = 1; \alpha_i \geq 0$

Want $q(x)$ a Gaussian

$$KL(p\|q) = \sum_{i=1}^{L}\alpha_i\left[\frac{1}{2}[\ln\det(S) + (\underline{\mu}_i - \underline{m})^T S^{-1}(\underline{\mu}_i - \underline{m}) + tr(S^{-1}\Sigma_i) - n]\right] + \frac{n}{2}\ln(2\pi) - H_p(\underline{x})$$

$$\nabla_{\underline{m}} KL(p\|q) = -\sum_{i=1}^{L}\alpha_i S^{-1}(\underline{\mu}_i - \underline{m}) = 0 \Rightarrow \underline{m} = \sum_{i=1}^{L}\alpha_i\,\underline{\mu}_i$$

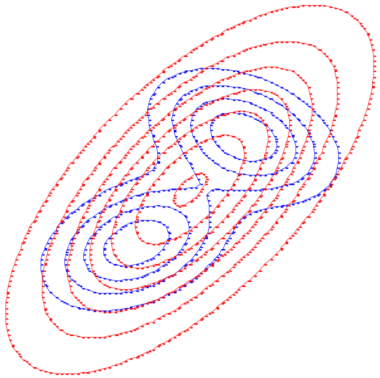
$$\nabla_S KL(p\|q) = \frac{1}{2}\sum_{i=1}^{L}\alpha_i[S^{-1} - S^{-1}(\underline{\mu}_i - \underline{m})(\underline{\mu}_i - \underline{m})^T S^{-1} - S^{-1}\Sigma_i S^{-1}] = 0$$

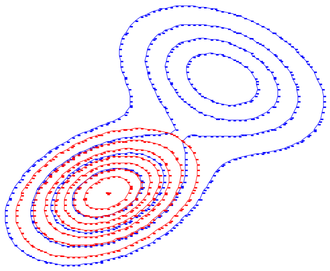
$\Rightarrow S = \sum_{i=1}^{L}\alpha_i[\Sigma_i + (\underline{\mu}_i - \underline{m})(\underline{\mu}_i - \underline{m})^T]$ ...M-projection, Moment projection, forward projection

$KL(q\|p)$ is messy because it involves ln of weighted sum of Gaussians!

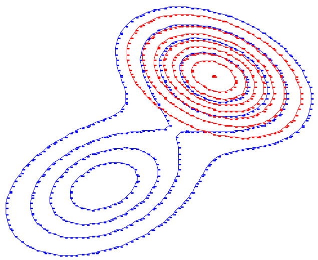
Need to do numerical optimization!..... I-projection, information projection, reverse projection

Generated using KLfwdReverseMixGauss from Murphy, Page 734



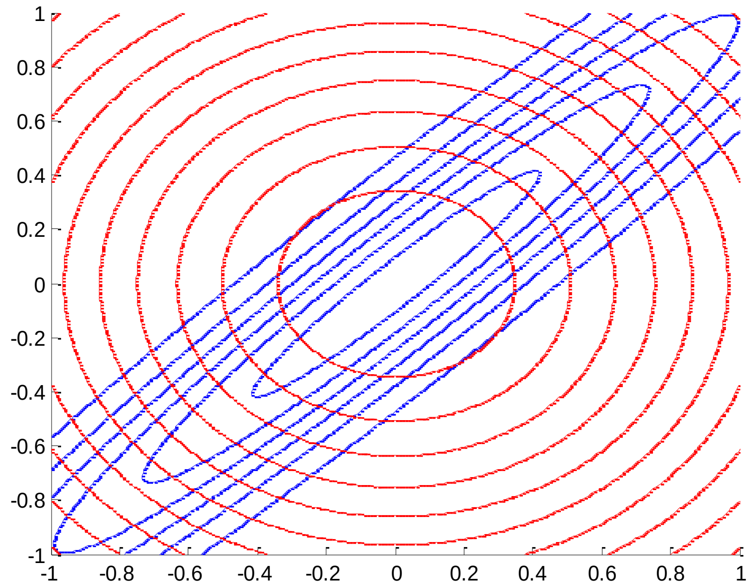


M-projection/Forward projection
Minimize KL($p||q$)/Match Moments
(overestimates uncertainty)

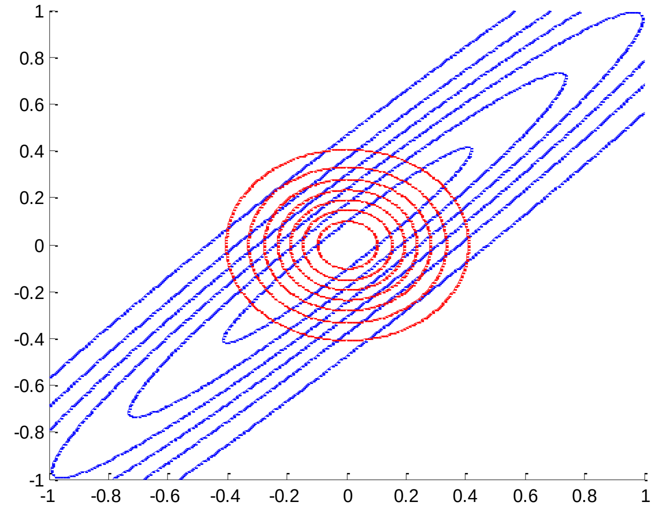


I-projection/Reverse projection
Minimize KL($q||p$)/Information projection
(too confident: Hit or Miss. Best keep both hypotheses)

Generated using KLpqGauss from Murphy, Page 734. *p* is correlated Gaussian and *q* is uncorrelated Gaussian



M-projection/Forward projection
Minimize KL(*p*||q)/Match Moments
(overestimates uncertainty)



I-projection/Reverse projection
Minimize KL(*q*||*p*)/Information projection
(underestimates uncertainty)

- Problem 6

$$p(x,y)=\begin{cases}1/3; x=0, y=0\\ 1/3; x=0, y=1\\ 0; x=1, y=0\\ 1/3; x=1, y=1\end{cases} \Rightarrow p(x)=\begin{cases}2/3; x=0\\ 1/3; x=1\end{cases} \text{ and } p(y)=\begin{cases}1/3; y=0\\ 2/3; y=1\end{cases}$$

$$p(y|x)=\frac{p(x,y)}{p(x)}=\begin{cases}1/2; x=0, y=0\\ 1/2; x=0, y=1\\ 0; x=1, y=0\\ 1; x=1, y=1\end{cases}; p(x|y)=\frac{p(x,y)}{p(y)}=\begin{cases}1; x=0, y=0\\ 0; x=1, y=0\\ 1/2; x=0, y=1\\ 1/2; x=1, y=1\end{cases}$$

$$H(x)=-(2/3)\log_2(2/3)-(1/3)\log_2(1/3)=0.9183\, bits=H(y)$$

$$H(y|x)=-\sum_{x=0}^{1}\sum_{y=0}^{1}p(x,y)\log_2 p(y|x)=0.6667\, bits=H(x|y)$$

$$H(x,y)=-\sum_{x=0}^{1}\sum_{y=0}^{1}p(x,y)\log_2 p(x,y)=1.585\, bits$$

(Note: $H(y|x)=H(x,y)-H(x)=1.585-0.9183=0.6667$)

- Problem 6

$$I(x,y) = H(x) - H(x|y) = 0.9183 - 0.6667 = 0.2516\, bits$$

$$(or) I(x,y) = KL(p(x,y) \| p(x)p(y)) = -\sum_{x=0}^{1}\sum_{y=0}^{1} p(x,y)\log_2 \frac{p(x)p(y)}{p(x,y)}$$

$$= -\frac{1}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{4}{3}) - \frac{1}{3}\log_2(\frac{2}{3}) = 0.2516\, bits$$

$$Note: H(x,y) = H(x) + H(y) - I(x,y) = H(x|y) + H(y|x) + I(x,y)$$

$$= H(x) + H(y/x) = H(y) + H(x|y)$$

Problem 7: Consider three Gaussian random variables x, y, z

$$p(x,y,z) = N(\begin{bmatrix}\mu_x \\ \mu_y \\ \mu_z\end{bmatrix}; \begin{bmatrix}\sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2\end{bmatrix}) = N(\begin{bmatrix}\mu_x \\ \mu_y \\ \mu_z\end{bmatrix}; \begin{bmatrix}\sigma_x^2 & \rho_{xy}\sigma_x\sigma_y & \rho_{xz}\sigma_x\sigma_z \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 & \rho_{yz}\sigma_y\sigma_z \\ \rho_{xz}\sigma_x\sigma_z & \rho_{yz}\sigma_y\sigma_z & \sigma_z^2\end{bmatrix})$$

Recall $\quad I(X;Y) = \frac{1}{2}[\ln|\Sigma_{xx}| - \ln|\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T|] = \frac{1}{2}[\ln|\Sigma_{yy}| - \ln|\Sigma_{yy} - \Sigma_{xy}^T\Sigma_{xx}^{-1}\Sigma_{xy}|]$

$x, y \; scalars : \Sigma_{xx} = \sigma_x^2; \Sigma_{xy} = \rho_{xy}\sigma_x\sigma_y; \Sigma_{yy} = \sigma_y^2$

$$\Rightarrow I(X;Y) = \frac{1}{2}[\ln\sigma_x^2 - \ln(\sigma_x^2(1-\rho_{xy}^2))] = -\frac{1}{2}\ln(1-\rho_{xy}^2)$$

Recall $I(X;Y;Z) = I(X;Y) - I(X;Y|Z) = -\frac{1}{2}\ln(1-\rho_{xy}^2) \; - I(X;Y|Z)$

Now, $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$

$$p(x|z) = N(\mu_x + \frac{\sigma_{xz}}{\sigma_z^2}(z-\mu_z); \sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2})$$

$$p(x|y,z) = N(\mu_x + \begin{bmatrix}\sigma_{xy} & \sigma_{xz}\end{bmatrix}\begin{bmatrix}\sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2\end{bmatrix}^{-1}\begin{bmatrix}(y-\mu_y) \\ (z-\mu_z)\end{bmatrix}; \sigma_x^2 - \begin{bmatrix}\sigma_{xy} & \sigma_{xz}\end{bmatrix}\begin{bmatrix}\sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{xy} \\ \sigma_{xz}\end{bmatrix})$$

$$I(X;Y|Z) = \frac{1}{2}[\ln(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2}) - \ln(\sigma_x^2 - \begin{bmatrix}\sigma_{xy} & \sigma_{xz}\end{bmatrix}\begin{bmatrix}\sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{xy} \\ \sigma_{xz}\end{bmatrix}_x^2)]$$

$$I(X;Y|Z)=\frac{1}{2}\ln\frac{(\sigma_x^2-\frac{\sigma_{xz}^2}{\sigma_z^2})(\sigma_y^2\sigma_z^2-\sigma_{yz}^2)}{\sigma_z^2(\sigma_x^2\sigma_y^2\sigma_z^2-\sigma_{yz}^2\sigma_x^2-\sigma_{xy}^2\sigma_z^2-\sigma_{xz}^2\sigma_y^2+2\sigma_{xy}\sigma_{yz}\sigma_{xz})}$$

$$=\frac{1}{2}\ln\frac{(1-\rho_{xz}^2)(1-\rho_{yz}^2)}{\underbrace{1-\rho_{xy}^2-\rho_{yz}^2-\rho_{xz}^2+2\rho_{xy}\rho_{xz}\rho_{yz}}_{\det er\min ant\ of\ 3x3\ correlation\ coeff.\ matrix\ \geq 0}}$$

$$\text{So, } I(X;Y;Z)=I(X;Y)-I(X;Y|Z)=-\frac{1}{2}\ln\frac{(1-\rho_{xy}^2)(1-\rho_{xz}^2)(1-\rho_{yz}^2)}{1-\rho_{xy}^2-\rho_{yz}^2-\rho_{xz}^2+2\rho_{xy}\rho_{xz}\rho_{yz}}$$

- Problem 9

$$J(\underline{w}) = \frac{1}{N}\sum_{n=1}^{N}(z_n - \underline{w}^T \underline{x}_n)^2 = \frac{1}{N}\|\underline{z} - X\underline{w}\|_2^2;\ \underline{z} = X\underline{w} + \underline{v}$$

$$(a)\nabla_{\underline{w}} J = \frac{2}{N} X^T\left(\underline{z} - X\underline{w}\right) = \underline{0} \Rightarrow \underline{\hat{w}} = (X^T X)^{-1} X^T \underline{z}$$

$$(b)\,\underline{\hat{z}} = X\underline{\hat{w}} = \underbrace{X(X^T X)^{-1} X^T}_{H}\underline{z} = Xw + H\underline{v}$$

$$(c)\,\underline{r} = \underline{z} - \underline{\hat{z}} = \underline{z} - X\underline{\hat{w}} = X\underline{w} + \underline{v} - Xw - H\underline{v} = (I_N - H)\underline{v}$$

$$(d)E[J(\underline{\hat{w}})] = \frac{1}{N}E\{\|\underline{z} - X\underline{\hat{w}}\|_2^2\} = \frac{1}{N}E\{\underline{v}^T(I_N - H)^2\underline{v}\}$$

$$= \frac{1}{N}E\{\underline{v}^T(I_N - H)\underline{v}\} = \frac{1}{N}trace[(I_N - H)E\{\underline{v}\underline{v}^T\}] = \sigma^2(1 - \frac{p}{N})$$

$$(e)z_{N+1} - \hat{z}_{N+1} = z_{N+1} - \underline{x}_{N+1}^T\underline{\hat{w}} = (v_{N+1} - \underline{x}_{N+1}^T(X^T X)^{-1} X^T \underline{v})$$

$$(f)E[(z_{N+1} - \hat{z}_{N+1})^2] = E[(v_{N+1} - \underline{x}_{N+1}^T(X^T X)^{-1} X^T \underline{v})^2]$$

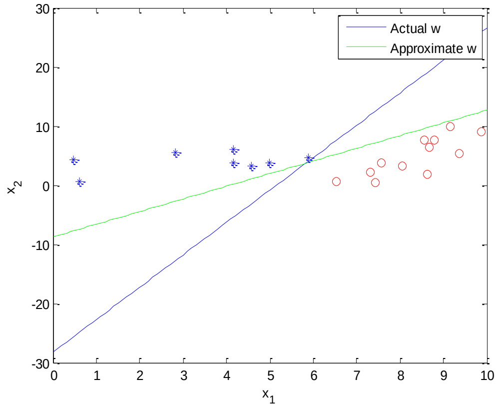
$$= E[v_{N+1}^2] - 2E[\underline{x}_{N+1}^T(X^T X)^{-1} X^T \underline{v} v_{N+1}] + E[\underline{x}_{N+1}^T(X^T X)^{-1} X^T \underline{v}\underline{v}^T X(X^T X)^{-1}\underline{x}_{N+1}]$$

$$= \sigma^2 + \sigma^2 trace\{(X^T X)^{-1} X^T X(X^T X)^{-1} E[\underline{x}_{N+1}\underline{x}_{N+1}^T]\} = \sigma^2 + \frac{\sigma^2}{N-1} trace\{\underbrace{(\frac{X^T X}{N-1})^{-1}}_{\approx\Sigma^{-1}}\Sigma\}$$

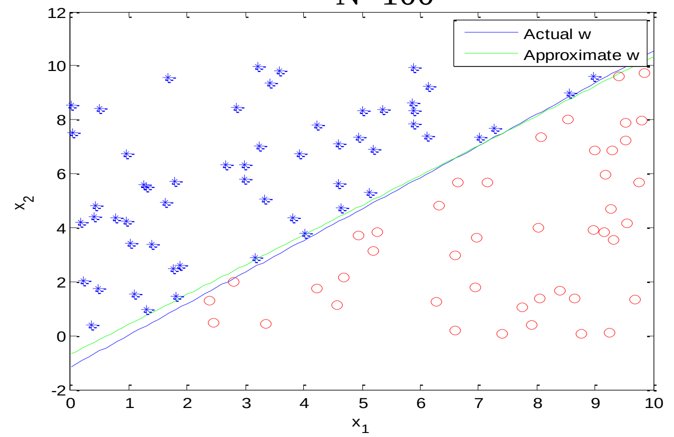
$$= \sigma^2(1 + \frac{p}{N-1}) \approx \sigma^2(1 + \frac{p}{N})$$

Problem 7

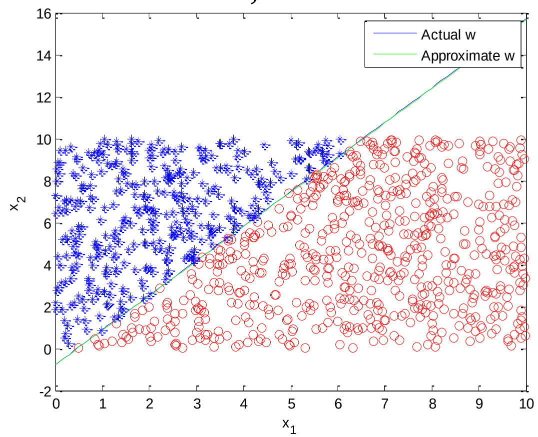
N=20



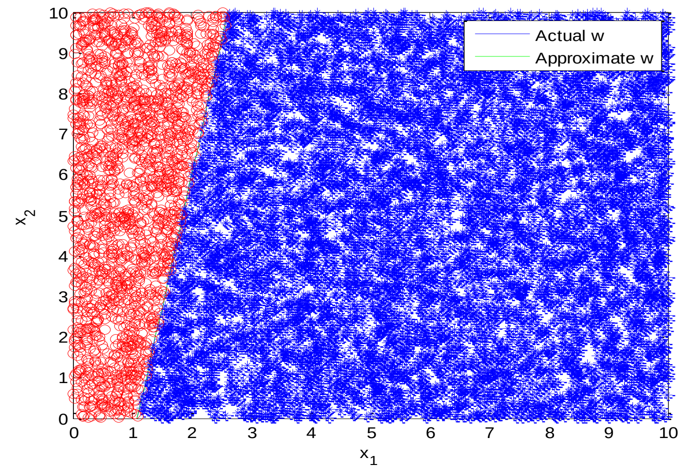
N=100



N=1,000



N=10,000



Problem 10

$$p(a,b,c) = p(a)\,p(b\,|\,a)\,p(c\,|\,b).$$

$$KL(q(a)\,p(b\,|\,a)\,\|\,p(a,b\,|\,c)) = \int_a\int_b q(a)p(b\,|\,a)\ln\frac{q(a)\,p(b\,|\,a)}{p(a,b\,|\,c)}\,dadb$$

$$Lagrangian: L(q(a),\lambda) = \int_a\int_b q(a)p(b\,|\,a)\ln\frac{q(a)\,p(b\,|\,a)}{p(a,b\,|\,c)}\,dadb + \lambda[\int_a q(a)da - 1]$$

$$= \int_a q(a)\ln q(a) + \int_a\int_b q(a)p(b\,|\,a)\ln\frac{p(b\,|\,a)}{p(a,b\,|\,c)}\,dadb + \lambda[\int_a q(a)da - 1]$$

$$\nabla_{q(a)}L = \ln q(a) + 1 + \int_b p(b\,|\,a)\ln\frac{p(b\,|\,a)}{p(a,b\,|\,c)}\,db + \lambda = 0$$

$$\Rightarrow \ln q(a) + \int_b p(b\,|\,a)\ln\frac{p(b\,|\,a)\,p(c)}{p(a)\,p(b\,|\,a)\,p(c\,|\,b)}\,db + 1 + \lambda = 0$$

$$\Rightarrow \ln\frac{q(a)}{p(a)} - \int_b p(b\,|\,a)\ln p(c\,|\,b)db + \underbrace{1+\lambda+\ln p(c)}_{constant\ indep.of\ a} = 0$$

$$q(a) \propto p(a)\exp[\int_b p(b\,|\,a)\ln p(c\,|\,b)db]$$

$$so, q(a) = \frac{p(a)\exp[\int_b p(b\,|\,a)\ln p(c\,|\,b)db]}{\int_a p(a)\exp[\int_b p(b\,|\,a)\ln p(c\,|\,b)db]da}$$

Problem 10

$$KL(q(a)p(b|a)\,\|\,p(a,b|c)) = \int_a\int_b q(a)p(b|a)\ln\frac{q(a)p(b|a)}{p(a,b|c)}dadb$$

$$q(a) = \frac{p(a)\exp[\int_b p(b|a)\ln p(c|b)db]}{\int_a p(a)\exp[\int_b p(b|a)\ln p(c|b)db]da} = \frac{p(a)f(c,a)}{Z(c)}$$

$$KL = \int_a\int_b q(a)p(b|a)[\ln q(a) + \ln p(b|a) - \ln p(a,b|c)]dadb$$

$$= \int_a\int_b q(a)p(b|a)[\ln q(a) + \ln p(b|a) - \ln p(a,b,c) + \ln p(c)]dadb$$

$$= \int_a\int_b q(a)p(b|a)[\ln q(a) + \ln p(b|a) - \ln p(a) - \ln p(b|a) - \ln p(c|b) + \ln p(c)]dadb$$

$$= \int_a\int_b q(a)p(b|a)[\ln q(a) - \ln p(a) + \ln p(c) - \ln p(c|b)]dadb$$

$$= \int_a q(a)\ln\frac{q(a)}{p(a)}da + \ln p(c) - \int_a q(a)\int_b p(b|a)\ln p(c|b)dbda$$

$$= \int_a q(a)\ln\frac{f(c,a)}{Z(c)}da + \ln p(c) - \int_a q(a)\ln f(c,a)da = \ln\frac{p(c)}{Z(c)}$$