Solution 3

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

$$p(\underline{z}) \sim N(\underline{z};\underline{\mu}_z,\Sigma_z);\ \underline{z}\in R^{n_z} \text{ and } p(\underline{x}\,|\,\underline{z}) \sim N(A\underline{z},\Sigma);\ \underline{x}\in R^{n_x}$$

$$\text{So, } p(\underline{x},\underline{z}) = N(A\underline{z},\Sigma)N(\underline{z};\underline{\mu}_z,\Sigma_z) \sim N\left(\begin{bmatrix}A\underline{\mu}_z\\ \underline{\mu}_z\end{bmatrix}, \begin{bmatrix}\overbrace{A\Sigma_z A^T+\Sigma}^{\Sigma_x} & \overbrace{A\Sigma_z}^{\Sigma_{xz}} \\ \underbrace{\Sigma_z A^T}_{\Sigma_{zx}} & \Sigma_z\end{bmatrix}\right)$$

$$E(\underline{z}\,|\,\underline{x}) = \underline{\mu}_z + \Sigma_{zx}\Sigma_x^{-1}(\underline{x}-A\underline{\mu}_z) = \underline{\mu}_z + \underbrace{\Sigma_z A^T\left[\overbrace{A\Sigma_z A^T+\Sigma}^{\text{Innovation covariance}}\right]^{-1}}_{\text{Kalman Gain}}(\underline{x}-A\underline{\mu}_z)$$

$$=[I_z\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A]\underline{\mu}_z + \Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}\underline{x}$$

Now consider

$$\left(\Sigma_z^{-1}+A^T\Sigma^{-1}A\right)^{-1} \overset{MIL}{=} [\Sigma_z\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A\Sigma_z] = [I_z\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A]\Sigma_z$$

$$\Rightarrow [I_z\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A] = \left(\Sigma_z^{-1}+A^T\Sigma^{-1}A\right)^{-1}\Sigma_z^{-1}$$

$$\left(\Sigma_z^{-1}+A^T\Sigma^{-1}A\right)^{-1}A^T\Sigma^{-1} = [\Sigma_z A^T\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A\Sigma_z A^T]\Sigma^{-1} = [\Sigma_z A^T\text{-}\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}(A\Sigma_z A^T+\Sigma-\Sigma)]\Sigma^{-1}$$

$$=\Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}$$

Then

$$E(\underline{z}\,|\,\underline{x}) = \left(\Sigma_z^{-1}+A^T\Sigma^{-1}A\right)^{-1}\left(A^T\Sigma^{-1}\underline{x}+\Sigma_z^{-1}\underline{\mu}_z\right)$$

$$\text{covar}(\underline{z}\,|\,\underline{x}) = \Sigma_z - \Sigma_{zx}\Sigma_x^{-1}\Sigma_{xz} = \Sigma_z - \Sigma_z A^T(A\Sigma_z A^T+\Sigma)^{-1}A\Sigma_z = \left(\Sigma_z^{-1}+A^T\Sigma^{-1}A\right)^{-1}$$

Problem 1

- Data: $p(\underline{x}_n \mid \underline{\mu}) = N(\underline{\mu}, \Sigma)$ *and* $p(\underline{\mu}) = N(\underline{\mu}_0, \Sigma_0)$

$$Let\ D^N = \{\underline{x}_1, \underline{x}_2, \ldots.., \underline{x}_{N-1}, \underline{x}_N\} = \{D^{N-1}, \underline{x}_N\}$$

$$p(D_N, \underline{\mu}) = p(\underline{x}_N \mid \underline{\mu}) p(D^{N-1} \mid \underline{\mu}) p(\underline{\mu}) = \prod_{n=1}^{N} p(\underline{x}_n \mid \underline{\mu}) = \left(\prod_{n=1}^{N} N(\underline{x}_n; \underline{\mu}, \Sigma) \right) N(\underline{\mu}; \underline{\mu}_0, \Sigma_0)$$

$$so,\ p(\underline{\mu} \mid D^N) = N(\underline{\mu}; \underline{\mu}_N, \Sigma_N) = N\left(\underline{\mu}; \Sigma_N \left(\Sigma_0^{-1} \underline{\mu}_0 + \Sigma^{-1} \sum_{n=1}^{N} \underline{x}_n \right), \Sigma_N \right)$$

$$where\ \Sigma_N = \left(\Sigma_0^{-1} + N\Sigma^{-1} \right)^{-1}$$

$$E(\underline{\mu} \mid D^N) = \Sigma_N \left(\Sigma_0^{-1} \underline{\mu}_0 + \Sigma^{-1} \sum_{n=1}^{N} \underline{x}_n \right)$$

$$\text{cov}(\underline{\mu} \mid D^N) = \Sigma_N = \left(\Sigma_0^{-1} + N\Sigma^{-1} \right)^{-1}$$

- Problem 2

$$LHS = \sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu})\ (\underline{x}_k^j - \underline{\mu})^T$$

$$= \sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu}_k + \underline{\mu}_k - \underline{\mu})\ (\underline{x}_k^j - \underline{\mu}_k + \underline{\mu}_k - \underline{\mu})^T$$

$$= \sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu}_k)\ (\underline{x}_k^j - \underline{\mu}_k)^T + \sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu}_k)(\underline{\mu}_k - \underline{\mu})^T +$$

$$\sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{\mu}_k - \underline{\mu})(\underline{x}_k^j - \underline{\mu}_k)^T + \sum_{k=1}^{c} n_k(\underline{\mu}_k - \underline{\mu})\ (\underline{\mu}_k - \underline{\mu})^T$$

$$= \sum_{k=1}^{c}\sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu}_k)\ (\underline{x}_k^j - \underline{\mu}_k)^T + \sum_{k=1}^{c} n_k(\underline{\mu}_k - \underline{\mu})\ (\underline{\mu}_k - \underline{\mu})^T \because \sum_{j=1}^{n_k}(\underline{x}_k^j - \underline{\mu}_k) = \underline{0}$$

$$= RHS$$

- Problem 3

$$a)\, p(D\,|\,\underline{\theta}) = p(\{\underline{x}_k\}_{k=1}^{n}\,|\,\underline{\theta}) = \prod_{k=1}^{n} p(\underline{x}_k\,|\,\underline{\theta})$$

$$= \prod_{k=1}^{n}\prod_{i=1}^{d} \theta_i^{x_{ik}}(1-\theta_i)^{1-x_{ik}}$$

$$= \prod_{i=1}^{d} \theta_i^{\sum_{k=1}^{n} x_{ik}}(1-\theta_i)^{n-\sum_{k=1}^{n} x_{ik}}$$

$$= \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i};\; s_i = \sum_{k=1}^{n} x_{ik}$$

$$b)\, p(\underline{\theta}\,|\,D) = \frac{p(D\,|\,\underline{\theta})\,p(\underline{\theta})}{p(D)} = \frac{p(D\,|\,\underline{\theta})\,p(\underline{\theta})}{\int_0^1 ..\int_0^1 p(D\,|\,\underline{\theta})\,p(\underline{\theta})d\underline{\theta}}$$

$$= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

- Problem 3

$$b)\, n = d = 1 \Rightarrow p(\theta \mid D) = 2\theta^{s}(1-\theta)^{1-s} = \begin{cases} 2(1-\theta);\, s=0 \\ 2\theta;\, s=1 \end{cases}$$

Easy to plot

$$c)\, p(\underline{x} \mid D) = \int_{\underline{\theta}} p(\underline{x} \mid \underline{\theta})\, p(\underline{\theta} \mid D)\, d\underline{\theta} = \int_{\underline{\theta}} \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i+x_i}(1-\theta_i)^{n+1-s_i-x_i}\, d\underline{\theta}$$

$$= \prod_{i=1}^{d} \frac{(s_i+x_i)!(n+1-s_i-x_i)!}{s_i!(n-s_i)!} \frac{1}{(n+2)}$$

$$\text{each term}: \begin{cases} \left(\dfrac{n+1-s_i}{n+2}\right) = 1 - \dfrac{s_i+1}{n+2};\, x_i = 0 \\ \left(\dfrac{s_i+1}{n+2}\right);\, x_i = 1 \end{cases}$$

$$\therefore\, p(\underline{x} \mid D) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}$$

$$d)\, \hat{\theta}_i = E(\theta_i \mid D) = \frac{s_i+1}{n+2};\, i = 1, 2, .., d$$

This will give you the same $p(\underline{x} \mid D) \Rightarrow \hat{\theta}_i$ *is a sufficient statistic*

- Problem 4

$$a)\ g_i(\underline{x}) = \sum_{k=1}^{N} \exp\left(\frac{\underline{w}_k^T \underline{x} - 1}{2\sigma^2}\right) a_{ki};\ a_{ki} = \begin{cases} 1\ if\ \ \underline{w}_k \in \omega_i \\ 0\ otherwsie \end{cases}$$

$N =$ total number of data points

$Since \parallel \underline{w}_k \parallel_2 = \parallel \underline{x} \parallel_2 = 1,\ g_i(\underline{x})\ can\ be\ written\ as$

$$g_i(\underline{x}) = \sum_{k=1}^{N} \exp\left(-\frac{\parallel \underline{x} - \underline{w}_k \parallel_2^2}{2\sigma^2}\right) a_{ki} \propto n_i \hat{p}(\underline{x} \mid \omega_i) \propto \hat{P}(\omega_i) \hat{p}(\underline{x} \mid \omega_i) \propto \hat{P}(\omega_i \mid \underline{x})$$

Here, n_i is the number of data points from class i. So, it does account for class priors.

b) The optimal classification rule for unequal costs is given by:

$$\hat{g}_i(\underline{x}) = \sum_{j=1}^{c} \lambda_{ij} \hat{P}(\omega_j) \hat{p}(\underline{x} \mid \omega_j) = \sum_{j=1}^{c} \lambda_{ij}\ g_j(\underline{x})$$

$\Rightarrow$ you can post-process PNNdiscriminants to take into account unequal costs.

- Problem 4

c) Training is unaffected because PNNjust normalizes patterns.

$$(i)\ compute\ g_i(\underline{x}) = \sum_{k=1}^{N} \exp\left(\frac{\underline{w}_k^T \underline{x} - 1}{2\sigma^2}\right) a_{ki};\ a_{ki} = \begin{cases} 1\ if\ \underline{w}_k \in \omega_i \\ 0\ otherwsie \end{cases}$$

$N =$ total number of data points

(ii)The optimal classification rule for unequal costs is given by:

$$\hat{g}_i(\underline{x}) = \sum_{j=1}^{c} \lambda_{ij}\ g_j(\underline{x}) \Rightarrow class = \arg\min \hat{g}_i(\underline{x})$$

Computation of $\hat{g}_i(\underline{x})$ can be embedded in PNN just as priors were.

Problem 5 (Problem 9.10 of Bishop)

$$a)\ p(\underline{x})=\sum_{k=1}^{K}\pi_k\, p(\underline{x}\,|\,k)$$

$$Let\ \underline{x}=\begin{bmatrix}\underline{x}_a\\ \underline{x}_b\end{bmatrix}$$

$$p(\underline{x}_b\,|\,\underline{x}_a)=\frac{p(\underline{x})}{p(\underline{x}_a)}=\frac{\sum_{k=1}^{K}\pi_k\, p(\underline{x}\,|\,k)}{\sum_{k=1}^{K}\pi_k\, p(\underline{x}_a\,|\,k)}=\sum_{k=1}^{K}\frac{\pi_k\, p(\underline{x}_a\,|\,k)}{\sum_{i=1}^{K}\pi_i\, p(\underline{x}_a\,|\,i)}p(\underline{x}_b\,|\,\underline{x}_a,k)$$

$$=\sum_{k=1}^{K}\hat{\pi}_k\, p(\underline{x}_b\,|\,\underline{x}_a,k);\ \hat{\pi}_k=\frac{\pi_k\, p(\underline{x}_a\,|\,k)}{\sum_{i=1}^{K}\pi_i\, p(\underline{x}_a\,|\,i)}=P(z=k\,|\,\underline{x}_a);z=latent\ \mathrm{var}\,iable$$

$$\Rightarrow Another\ mixture\ density$$

$$Note: For\ Gaussian\ p(\underline{x}\,|\,k)=N(\underline{\mu}_k,\Sigma_k)$$

$$where\ \underline{\mu}_k=\begin{pmatrix}\underline{\mu}_{a,k}\\ \underline{\mu}_{b,k}\end{pmatrix},\Sigma_k=\begin{pmatrix}\Sigma_{aa,k} & \Sigma_{ab,k}\\ \Sigma_{ba,k} & \Sigma_{bb,k}\end{pmatrix}, we\ have$$

$$p(\underline{x}_a\,|\,k)=N(\underline{\mu}_{a,k},\Sigma_{aa,k})$$

$$p(\underline{x}_b\,|\,\underline{x}_a,k)=N(\underline{\mu}_{b,k}+\Sigma_{ba,k}\left(\Sigma_{aa,k}\right)^{-1}(\underline{x}_a-\underline{\mu}_{a,k}),\Sigma_{bb,k}-\Sigma_{ba,k}\left(\Sigma_{aa,k}\right)^{-1}\Sigma_{ab,k})$$

- Problem 6 (Problem 9.19 of Bishop)

$$a)\, p(X,Z\,|\,\{\mu_{kij}\}_{k=1,i=1,j=1}^{K\ \ D\ \ M};\{\pi_k\}_{k=1}^{K}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\pi_k^{z_{nk}}[\prod_{i=1}^{D}\prod_{j=1}^{M}\mu_{kij}{}^{x_{nij}}]^{z_{nk}};\, z_{nk}\in(0,1);\, \mu_{kij} \ni \sum_{j=1}^{M}\mu_{kij}=1\forall k,i$$

$$x_{nij}\in(0,1)\ni\sum_{j=1}^{M}x_{nij}=1\ \forall n,i;\sum_{k=1}^{K}\pi_k=1;\pi_k\ge 0$$

$$\Rightarrow \ln p(X,Z\,|\,\{\mu_{kij}\}_{k=1,i=1,j=1}^{K\ \ D\ \ M};\{\pi_k\}_{k=1}^{K}) = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{D}\sum_{j=1}^{M} z_{nk}[\ln\pi_k + x_{nij}\ln\mu_{kij}]$$

$$E-step: P(z_{nk}=1\,|\,X,\{\mu_{kij}^{old}\}_{k=1,i=1,j=1}^{K\ \ D\ \ M};\{\pi_k^{old}\}_{k=1}^{K}) = E(z_{nk}) = \gamma(z_{nk}) = \frac{\pi_k^{old}\prod_{i=1}^{D}\prod_{j=1}^{M}\mu_{kij}{}^{x_{nij}}}{\sum_{l=1}^{K}\pi_l^{old}\prod_{i=1}^{D}\prod_{j=1}^{M}\mu_{lij}{}^{x_{nij}}};\, n=1,2,..,N$$

$$M-step: N_k = \sum_{n=1}^{N}\gamma(z_{nk})\ \forall k$$

$$\pi_k^{new} = \frac{N_k}{N}\ \forall k$$

$$\mu_{kij} = \frac{\sum_{n=1}^{N}\gamma(z_{nk})x_{nij}}{N_k}\ \forall i,j$$

- Problem 7 (Problem 10.12 of Bishop)

$$p(\{\underline{x}^n\}_{n=1}^N,\{\underline{z}^n\}_{n=1}^N,\{P_j,\underline{\mu}_j,\Sigma_j\}_{j=1}^M) = p(\{\underline{x}^n\}_{n=1}^N \mid \{\underline{z}^n\}_{n=1}^N,\{\underline{\mu}_j,\Sigma_j\}_{j=1}^M).$$

$$p(\{\underline{z}^n\}_{n=1}^N \mid \{P_j\}_{j=1}^M).\ p(\underline{P}).p(\{\underline{\mu}_j\}_{j=1}^M \mid \{\Sigma_j\}_{j=1}^M).p(\{\Sigma_j\}_{j=1}^M)$$

$$=\left(\prod_{n=1}^N\prod_{j=1}^M [P_j N(\underline{x}^n;\underline{\mu}_j,\Sigma_j)]^{z_j^n}\right).$$

$$\frac{\Gamma(M\alpha_0)}{(\Gamma(\alpha_0))^M}\left(\prod_{j=1}^M P_j^{\alpha_0-1}.N(\underline{\mu}_j;\underline{m}_0,\frac{1}{\beta_0}\Sigma_j).\ Wishart(\Sigma_j^{-1};\nu_0,W_0)\right)$$

$$\ln q(\{\underline{z}^n\}_{n=1}^N) = E_{q(\{P_j,\underline{\mu}_j,\Sigma_j\}_{j=1}^M)}\left(\ln p(\{\underline{x}^n\}_{n=1}^N,\{\underline{z}^n\}_{n=1}^N,\{P_j,\underline{\mu}_j,\Sigma_j\}_{j=1}^M)\right)$$

$$=E_{q(\{P_j,\underline{\mu}_j,\Sigma_j\}_{j=1}^M)}\left(\ln\prod_{n=1}^N\prod_{j=1}^M [P_j N(\underline{x}^n;\underline{\mu}_j,\Sigma_j)]^{z_j^n}\right)+cons\tan t$$

$$=E_{q(\{\underline{\mu}_j,\Sigma_j\}_{j=1}^M)}\left(\ln p(\{\underline{x}^n\}_{n=1}^N \mid \{\underline{z}^n\}_{n=1}^N,\{\underline{\mu}_j,\Sigma_j\}_{j=1}^M)\right)+$$

$$E_{q\{\underline{P}\}}\left(\ln p(\{\underline{z}^n\}_{n=1}^N \mid \{P_j\}_{j=1}^M)\right)+cons\tan t$$

$$=\sum_{n=1}^N\sum_{j=1}^M z_j^n \ln\rho_j^n$$

where $\ln\rho_j^n = E_{P_j}\left[\ln P_j\right]+\frac{1}{2}E_{\Sigma_j^{-1}}\left[\ln|\Sigma_j^{-1}|\right]-\frac{p}{2}\ln(2\pi)-\frac{1}{2}E_{\underline{\mu}_j,\Sigma_j^{-1}}\left(\|\underline{x}^n-\underline{\mu}_j\|_{\Sigma_j^{-1}}^2\right)$

$\Rightarrow q(\{\underline{z}^n\}_{n=1}^N) = \prod_{n=1}^N\prod_{j=1}^M [\gamma_j^n]^{z_j^n}$ where $\gamma_j^n = \dfrac{\rho_j^n}{\sum_{k=1}^M \rho_k^n}$ .... responsibilities

- Problem 8 (Problem 10.14 of Bishop)

$$E_{\underline{\mu}_j,\Sigma_j^{-1}}\left(\|\underline{x}^n-\underline{\mu}_j\|^2_{\Sigma_j^{-1}}\right)=\int_{\Sigma_j^{-1}}\int_{\underline{\mu}_j}\|\underline{x}^n-\underline{\mu}_j\|^2_{\Sigma_j^{-1}}\; N(\underline{\mu}_j;\underline{m}_j,\frac{1}{\beta_j}\Sigma_j).Wishart(\Sigma_j^{-1};\nu_k,W_k)d\underline{\mu}_j d\Sigma_j^{-1}$$

$$Inner\ Integral:tr\left(\Sigma_j^{-1}\underline{x}^n\left(\underline{x}^n\right)^T-\Sigma_j^{-1}\underline{x}^n\underline{m}_j^T-\Sigma_j^{-1}\underline{m}_j\left(\underline{x}^n\right)^T+\Sigma_j^{-1}\underline{m}_j\underline{m}_j^T+\frac{1}{\beta_j}I_p\right)$$

$$For\ Wishart,\ E(\Sigma_j^{-1})=\nu_k W_k$$

So, the outer integral is:

$$E_{\underline{\mu}_j,\Sigma_j^{-1}}\left(\|\underline{x}^n-\underline{\mu}_j\|^2_{\Sigma_j^{-1}}\right)=tr\left(\nu_k W_k\left(\underline{x}^n\left(\underline{x}^n\right)^T-\underline{x}^n\underline{m}_j^T-\underline{m}_j\left(\underline{x}^n\right)^T+\underline{m}_j\underline{m}_j^T\right)+\frac{1}{\beta_j}I_p\right)$$

$$=\frac{p}{\beta_j}+\nu_k\|\underline{x}^n-\underline{m}_j\|^2_{W_k}$$