Worksheet 3

Problem 1. General view of GMM [Bi] Ex. 9.9

Recall that The expected value of the complete-data log likelihood function for GMM is given by

$$\mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\right] = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\left\{\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right\}.$$

With a fixed $\gamma(z_{nk})$, find the maximizer $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ for $\mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\right]$.

Problem 2. K-means as the limit of EM cf. [Bi] Ch.9.3.2

Consider the EM algorithm where the covariance matrices of the mixture components are all given by $\boldsymbol{\Sigma}_k = \epsilon \boldsymbol{I},\ k = 1, \cdots, K.$

1. Write $p(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.
2. Show that $\gamma(z_{nk}) \to r_{nk}$ as $\epsilon \to 0$, where $r_{nk} = 1$ if $k = \operatorname{argmin}_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2$ and $r_{nk} = 0$ otherwise.
3. Show that as $\epsilon \to 0$,

$$\mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\right] \to -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\left\|\mathbf{x}_n - \boldsymbol{\mu}_k\right\|^2 + \text{const}.$$

Problem 3. Rayleigh quotient

The Rayleigh quotient for a real symmetric matrix $\boldsymbol{A}$ and a nonzero vector $\boldsymbol{v}$ is given by

$$\rho(\boldsymbol{v}, \boldsymbol{A}) = \frac{\boldsymbol{v}^\mathsf{T}\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^\mathsf{T}\boldsymbol{v}}.$$

Prove that the $\rho(\boldsymbol{v}, \boldsymbol{A}) \in [\lambda_{\min}, \lambda_{\max}]$ where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of $\boldsymbol{A}$, respectively. For what $\boldsymbol{v}$ does $\rho(\boldsymbol{v}, \boldsymbol{A})$ achieve the min and the max, respectively?

Problem 4. Graph Laplacian

1. Prove that all the eigenvalues of the graph Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ are non-negative.
2. Prove that all the eigenvalues of the normalized graph Laplacian $\boldsymbol{L}_{\text{sym}} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$ are in $[0, 2]$.

Problem 5. One-class SVM

The optimization problem for one-class SVM is

$$\begin{aligned}&\min\ R^2 + C\sum_{n=1}^{N}\xi_n\\ &\text{s.t. } \|\phi(\boldsymbol{x}_n)-\boldsymbol{a}\|^2 \le R^2+\xi_n \text{ for all } n\\ &\qquad \xi_n \ge 0 \text{ for all } n\end{aligned}$$

Write the Lagrangian and express it using only the Lagrange multipliers and the kernel $K(\boldsymbol{x}_n, \boldsymbol{x}_m) = \phi(\boldsymbol{x}_n)^\mathsf{T}\phi(\boldsymbol{x}_m)$.

Problem 6. RKHS cf. [HaTF] Ex.5.16

Recall that $K(x,y) = \sum_{j=1}^{\infty}\gamma_j\phi_j(x)\phi_j(y)$ for which we can order $\gamma_1 \ge \gamma_2 \ge \cdots$ and $\{\phi_j\}_{j=1}^{\infty}$ is orthonormal: $\langle\phi_i,\phi_j\rangle = \delta_{ij}$. Consider the ridge regression problem

$$\min_{\{c_j\}_{j=1}^{\infty}} \sum_{n=1}^{N}\left(y_n - \sum_{j=1}^{\infty}c_j\phi_j(x_n)\right)^2 + \lambda\sum_{j=1}^{\infty}\frac{c_j^2}{\gamma_j}\ ,$$

1. Explain why the problem is equivalent to

$$\min_{\boldsymbol{\alpha}}(\boldsymbol{y}-\boldsymbol{K\alpha})^\mathsf{T}(\boldsymbol{y}-\boldsymbol{K\alpha}) + \lambda\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{K\alpha}\ .$$

2. Assume $K(x,y) = \sum_{m=1}^{M}h_m(x)h_m(y)$ and $M \ge N$. Prove:

$$\boldsymbol{h}(x) = \boldsymbol{V}\boldsymbol{D}_\gamma^{1/2}\boldsymbol{\phi}(x)$$

where $\boldsymbol{h}(x) = [h_1(x),\cdots,h_M(x)]^\mathsf{T}$ and $\boldsymbol{\phi}(x) = [\phi_1(x),\cdots,\phi_M(x)]^\mathsf{T}$; $\boldsymbol{V}$ is an $M\times M$ orthogonal matrix and $\boldsymbol{D}_\gamma = \text{diag}(\gamma_1,\cdots,\gamma_M)$. What are $\boldsymbol{V}$ and $\boldsymbol{D}_\gamma$? (Hint: $h_m = \sum_{j=1}^{M}\langle h_m,\phi_j\rangle\phi_j$).