Problem Set 1

March 3, 2017

Problem 1

(a) Consider a real, symmetric matrix Σ whose eigenvalue equation is given by:

$$\Sigma u_i = \lambda_i u_i$$

By taking the complex conjugate of this equation and subtracting the original equation, and then forming the inner product with eigenvector u_i, show that the eigenvalues λ_i are real. Similarly, use the symmetry property of Σ to show that two eigenvectors u_i and u_j will be orthogonal provided $\lambda_j \neq \lambda_i$. Finally, show that without loss of generality, the set of eigenvectors can be chosen to be orthonormal, even if some of the eigenvalues are zero.
(b) Refer to slides about PCA, where we perform eigen-decomposition on

$$A = \frac{1}{N}\sum_{i=1}^{N} x_i {x_i}^\top$$

Prove A is a symmetric and positive semi-definite matrix.

Problem 2

Given a set of i.i.d. data $X = \{x_1, \cdots, x_N\}$ drawn from $N(x; \mu, \Sigma)$, we want to estimate (μ, Σ) by MLE.
(a) Write the log likelihood function.
(b) Take the derivative of log likelihood function w.r.t. μ, show that

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

.
(c) Take the derivative of log likelihood function w.r.t. Σ, show that

$$\Sigma_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_{\mathrm{ML}})(x_i - \mu_{\mathrm{ML}})^\top$$

The following derivatives (or refer to The Matrix CookBook[1]) may be helpful (X, B are matrices, x, a, b are vectors):

$$\begin{aligned}
\frac{\partial \log(|\det X|)}{\partial X} &= \left(X^{-1}\right)^\top = \left(X^\top\right)^{-1} \\
\frac{\partial a^\top X^{-1} b}{\partial X} &= -X^{-\top} a b^\top X^{-\top} \\
\frac{\partial x^\top B x}{\partial x} &= (B + B^\top)x
\end{aligned} \tag{1}$$

(d) Evaluate expectations of μ_{ML} and Σ_{ML}, show μ_{ML} is unbiased but $\Sigma_{\mathrm{ML}}$is biased.

Problem 3

For support vector machines, the class-conditional distributions may overlap, we therefore modify the support vector machine so as to allow some of the training points to be misclassified. For un-separable case, the formalization of the optimal problem becomes: Given $\{x_i, y_i\}, i = 1, ..., N, y_i \in \{-1, 1\}$ are training examples,

$$\begin{aligned}
&\min_{w,b} \frac{||w||^2}{2} + C \sum_{i=1}^{N} \xi_i \\
&\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i,\ 1 \leq i \leq N \\
&\qquad \xi_i \geq 0, 1 \leq i \leq N
\end{aligned} \tag{2}$$

where the ϵ_i denotes the slack variable penalty, and the parameter C controls the trade-off betwwen the slack variable penalty and the margin.

Please give the solutions of w and b.

(a) Give the corresponding Lagrangian and the set of KKT conditions.

(b) Optimize out w, b and $\{\epsilon_i\}$.

(c) Give the dual Lagrangian.

(d) Give the final solution for w and the numerically stable solution of b.

[1] www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf