Homework I: DATA620013

Advanced Statistical Learning

Due Tuesday, Nov 5th, 2020

1 Problem 1

Prove Gauss Markov Theorem: suppose $\hat{\beta}$ is the OLS estimator. To estimate $c^\top \beta$, consider a linear estimator $c^\top \hat{\beta} = l^\top y$ that is unbiased for $c^\top \beta$, and an arbitrary unbiased estimator $d^\top y$ (with $d \neq l$), show that $var(d^\top y) > var(l^\top y)$

2 Problem 2

Assume a K class problem where $f(x \mid G = k) \sim \mathcal{N}(\mu_k, \Sigma)$, $k = 1, ..., K$. (The covariance is the same in all classes). Assume $Pr(G = k) = \pi_k$, $k = 1..., K$. Let $(X_1, G_1), (X_2, G_2), ..., (X_N, G_N)$ be an i.i.d sample from the joint distribution.

Write the likelihood of the data. Show that $\hat{\mu}_k$ (the maximum likelihood estimator of μ_k) is given by the sample means for the data for which $G_n = k$. Show that the maximum likelihood estimate $\hat{\Sigma}$ is given by the pooled covariance estimate, i.e. a weighted average of the separate sample covariances for each class, and provide the weights.

3 Problem 3

Suppose x is m-dim random variable with covariance matrix Σ, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0$ are the eigenvalues of Σ, and $\alpha_1, \alpha_2, ..., \alpha_m$ are the corresponding eigenvectors, show that the k-th principle component of x is given by

$$y_k = \alpha_k^\top x = \alpha_{1k} x_1 + \alpha_{2k} x_2 + ... + \alpha_{mk} x_m$$

for $k = 1, 2, ..., m$, and the variance of y_k is given by (the k-th eigenvalue of Σ)

$$var(y_k) = \alpha_k^\top \Sigma \alpha_k = \lambda_k$$

4 Problem 4

In the setting of logistic regression, we assume $Pr(G = 1 \mid X = x, \beta) = [1 + e^{-\beta^\top x}]^{-1}$ and there are i.i.d data $x_1, ..., x_n \in \mathbb{R}^p$ and $g_1, ..., g_n \in \{0, 1\}$. We now add a prior $\beta \sim \mathcal{N}_p(0, \alpha^2 I)$. We are interested in the $\hat{\beta}$ that maximize the posterior of β.

Write $\hat{\beta}$ as the solution of an empirical loss minimization problem and give the explicit expression of the loss function. Note that you need to consider the different way of coding y_i by $t_i = 2y_1 - 1 \in \{-1, 1\}$