MATH437/537 Fall, 2022

Homework 7. Due December 5

1. (Sample linear discriminant analysis) Let $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ be the following 14×1 vectors:

$$\boldsymbol{\mu}_1 = \mathbf{0}, \quad \boldsymbol{\mu}_2 = (3, \ldots, 3)^\top, \quad \boldsymbol{\mu}_3 = (-3, \ldots, -3)^\top.$$

   (a) Draw a sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{16}$ iid $N_{14}(\boldsymbol{\mu}_1, 4\boldsymbol{I})$, a sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_8$ iid $N_{14}(\boldsymbol{\mu}_2, 4\boldsymbol{I})$ and a sample $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_8$ iid $N_{14}(\boldsymbol{\mu}_3, 4\boldsymbol{I})$. These samples give you a data matrix for each of three classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$.

   (b) Write the formulas you will use to compute $\widehat{\boldsymbol{B}}_\mu$ and $\boldsymbol{S}_{\text{pooled}}$.

   (c) The kth sample discriminant of a point $\boldsymbol{x}$ is $\boldsymbol{y}_k = \boldsymbol{w}_k^\top \boldsymbol{x}$. How is $\boldsymbol{w}_k$ defined?

   (d) Make a plot of all the data on the first two discriminants using different colors for each of the three classes.

2. Consider a linear regression problem where $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\boldsymbol{X}$ is a 100×10 matrix with independent columns, and $\boldsymbol{\varepsilon} \sim N_{100}(\mathbf{0}, \sigma^2\boldsymbol{I})$.

   (a) The fitted vector $\widehat{\boldsymbol{y}}$ is the projection of $\boldsymbol{y}$ onto what subspace of $\mathbb{R}^{100}$?

   (b) Let $\boldsymbol{P}$ be the projection matrix such that $\widehat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$. What is the trace of $\boldsymbol{P}$?

   (c) The residuals of the fit are $\boldsymbol{r} = \boldsymbol{Q}\boldsymbol{y}$ where $\boldsymbol{Q}$ is an orthogonal projection onto what subspace of $\mathbb{R}^{100}$?

   (d) What is the trace of $\boldsymbol{Q}$?

   (e) Determine $\mathbb{E}\,\|\boldsymbol{r}\|^2 = \mathbb{E}\,(\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y})$. (Hint: Note that $\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{Q}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ and use a formula given in class.)

   (f) Find an unbiased estimator of σ^2.