CS 559: Midterm
Duration: 2:00

Answer all problems. Show all calculations and provide sufficient explanation. If in doubt, explain more.

Write your name on this page and additional pages (if necessary)

NAME:

Problem 1. (10 Points) In the context of Fisher's Linear Discriminant Analysis, explain why maximizing the distance between the projected class means is not sufficient for obtaining well-separated data after the projection. A simple sketch may be helpful for your answer.

Problem 2. (15 Points)

(1) Maximum Likelihood Estimation (MLE) techniques assume a certain parametric form for the class-conditional probability density functions. This implies that (select one only) (5 pt):

a. The form of the decision boundaries is also determined in some cases.

b. The form of the decision boundaries is always unpredictable.

(2) Linear Discriminant Analysis techniques assume a certain parametric form for the decision boundaries. This implies that (select one only) (5 pt):

a. The form of the class-conditional densities is also determined in some cases.

b. The form of the class-conditional densities remains unknown in general.

(3) Briefly explain your above-two answers. (5 pts)

Problem 3. (15 points) Let D denote the data samples and H denote hypothesis. Provide the relationship between the following probability pairs, using one of the following operators:
$(1) =, (2) \leq, (3) \geq$, and (4) (depends)
Explain your answers briefly.

(a) $\sum_h P(H = h|D = d)$ and 1 (3 pts)

(b) $\sum_h P(D = d|H = h)$ and 1 (3 pts)

(c) $\sum_h P(D = d|H = h)P(H = h)$ and 1 (3 pts)

(d) $P(H = h|D = d)$ and $P(H = h)$ (3 pts)

(e) $P(H = h|D = d)$ and $P(D = d|H = h)P(H = h)$ (3 pts)

Problem 4. (20 points) Let $\mathbf{x}$ be a one-dimensional binary (0 or 1) variable following a Bernoulli distribution:

$$P(\mathbf{x}|\theta) = \theta^x(1-\theta)^{1-x},$$

where θ, the probability that $x = 1$, is the unknown parameter to be estimated. Show that the maximum-likelihood estimate for θ is

$$\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

Problem 5. True or False. No explanation needed. (20 points)

(1) MLE and MAP never produce the same result.

(2) Posterior is always higher than prior.

(3) One can find a closed-form solution for any optimziation problem.

(4) When applying generative models, we usually assume that the parameter estimation for each class is independent.

(5) K-NN is a parameteric approach.

(6) Histogram estimation is a parametric approach.

(6) A belief network is a directed acyclic graph.

(7) PCA can be sovled using SVD.

(8) PCA can be applied for face detection.

(9) Covariance matrix captures the shape of a distribution.

(10) A susbpace must pass through the origin.

Problem 6. (20 points) Assume we are given a set of D dimensional data samples.
PCA: Principal Component Analysis LDA: Fisher Linear Discriminant Analysis

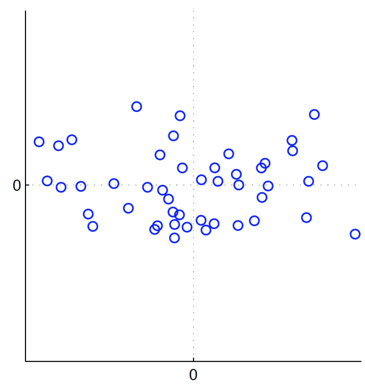
(a) Which quantity does PCA maximize in order to obtain the first projection direction? (3 pts)

(b) Which quantity does PCA minimize in order to obtain the first projection direction? (3 pts)

(c) Which quantity does LDA maximize in order to obtain the first projection direction? (3 pts)

(d) Consider a data set with two data points: $(2, 2), (-2, -2)$. Compute the covariance matrix (3 pts)

Suppose the covariance matrix of the two-dimensional data set plotted below is $\begin{bmatrix} \alpha, 0 \\ 0, \beta \end{bmatrix}$. We assume the horizontal axis corresponds to the first dimension and the vertical one corresponds to the second.



(e) Draw the first principal direction estimated from the data. (4 pts)

(f) How large is the variance of the data projected on the first principal component? (4 pts)