STATS 261: Homework 1

Due April 15 at 3pm.

(Submit electronically on CCLE if able; else, email to ruiqigao@ucla.edu. Do not bring to class.)

1. *Supervised vs. unsupervised learning.* For each problem below, state whether the task is best posed as a supervised or unsupervised learning problem. Explain your answer.

   (a) Finding the distribution of values of a set of heat sensors to detect outliers.

   (b) Predicting if a patient has cancer based on an medical image from the patient.

   (c) Grouping reviews together that have similar topics.

2. *Selecting targets and predictors.* Suppose users browse products on an online company's website. The company wants to predict what product a visitor to the website is most interested in seeing next.

   (a) To set this up as a supervised learning problem, what variables would you select as the predictors and what variables would be the targets?

   (b) Where would they get the data to train the model?

   There is no single correct answer. Try to be as detailed as possible in your description of the predictors and targets. State whatever assumptions you need to make, such as the format of the product.

3. *Bayes rule with a binary random variable.* In some population, a disease occurs with probability $p_0 = 10^{-3}$. A test for the disease has the following properties:

   - If a person is sick, the test is positive with probability 90%.
   - If a person is healthy, the test is negative with probability 80%.

   Find the probability that the person is sick given a positive test result and the probability that they are sick given a negative result.

4. *Bayes rule.* A person has five coins: three fair coins, one double-headed coin and one double-tailed coin. He selects one of the five coins at random and tosses it.

   (a) What is the probability that the top of the coin that was tossed is a head?

   (b) What is the probability that the bottom of the coin is a head, given that the top of the coin is a head?

5. *Joint and conditional densities.* Consider the joint density $f_{X,Y}(x, y) = ce^{-xy}$, $x \geq 0, y \in [1, 2]$.

   (a) Find the constant c.

(b) Find the marginal density $f_X(x)$.

(c) Find the conditional density $f_{Y|X}(y|x)$.

6. *Scalar classification.* Suppose that we are trying to estimate a binary class label $y = 0$ or 1 from a scalar random variable $x \in [-1, 1]$ with likelihood functions

$$p(x|y=0) = \frac{1}{2}, \qquad p(x|y=1) = \frac{3}{2}x^2. \tag{1}$$

(a) Find the maximum likelihood (ML) classifier, $\widehat{y}$, in terms of x.

(b) For the ML classifier, compute the probabilities of missed detection and false alarm:

$$P_{\rm MD} = \Pr(\widehat{y} = 0|y = 1), \qquad P_{\rm FA} = \Pr(\widehat{y} = 1|y = 0). \tag{2}$$

(c) Let $L(\widehat{y}, y)$ be the loss function

$$L(\widehat{y}, y) = \begin{cases} c_1 & \text{if } \widehat{y} = 1, y = 0, \\ c_2 & \text{if } \widehat{y} = 0, y = 1, \\ 0 & \text{if } \widehat{y} = y, \end{cases} \tag{3}$$

where c_1 and c_2 are the costs of false alarms and missed detections, respectively. Suppose that

$$P(y = 1) = q, \quad P(y = 0) = 1 - q,$$

for some q. For the ML detector in part (a), what is the expected Bayes risk $\mathbb{E}(L(\widehat{y}, y))$?

7. *ROC curve and risk minimization.* Consider binary classification from two scalar exponentials:

$$p(x|y = j) = \lambda_j e^{-\lambda_j x}, \quad x \geq 0, \quad j = 0, 1. \tag{4}$$

Assume $\lambda_0 > \lambda_1$. Let $\widehat{y}$ be the classifier:

$$\widehat{y} = \begin{cases} 1 & \text{if } x \geq t, \\ 0 & \text{if } x \leq t, \end{cases}$$

where t is a threshold.

(a) Compute the missed detection probability $P_{\rm MD}(t)$ and false alarm probability $P_{\rm FA}(t)$ as functions of t.

(b) For the remainder of the problem, let $\lambda_0 = 1$ and $\lambda_1 = 5$. Plot the ROC curve, $P_{\rm D}(t) = 1 - P_{\rm MD}(t)$ vs. $P_{\rm FA}(t)$.

(c) Suppose that missed detections are ten times as costly as false alarms and both classes are equiprobable. What is the optimal threshold t?

8. *Vector classifiers.* Suppose that we are trying to estimate a binary class label $y = 0$ or 1 from a vector random variable $\mathbf{x} \in \mathbb{R}^d$ with conditional Gaussian likelihoods:

$$p(\mathbf{x}|y = j) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \mathbf{S}_j), \tag{5}$$

where $\boldsymbol{\mu}_j \in \mathbb{R}^d$ and $\mathbf{S}_j \in \mathbb{R}^{d\times d}$ are the mean and covariance matrices in each class. Let $\widehat{y}$ be the LRT classifier,

$$\widehat{y} = \begin{cases} 1 & \text{if } L(\mathbf{x}) \geq \gamma, \\ 0 & \text{if } L(\mathbf{x}) < \gamma, \end{cases} \quad L(\mathbf{x}) = \ln \left[\frac{p(\mathbf{x}|y=1)}{p(\mathbf{x}|y=0)} \right], \tag{6}$$

where γ is a threshold.

(a) Write the log likelihood ratio $L(\mathbf{x})$ in terms of $\mathbf{x}$, $\boldsymbol{\mu}_j$ and $\mathbf{S}_j$. To refresh your memory, you may wish to look up the formula for the density of a multivariable Gaussian.

(b) Suppose that $\mathbf{S}_0 = \mathbf{S}_1 = \sigma^2 \mathbf{I}$. Show that the classifier (8) can be written using a *linear discriminant*. That is,

$$\widehat{y} = \begin{cases} 1 & \text{if } g(\mathbf{x}) \geq t, \\ 0 & \text{if } g(\mathbf{x}) < t, \end{cases} \quad g(\mathbf{x}) = \mathbf{c}^\intercal \mathbf{x}. \tag{7}$$

Find $\mathbf{c}$ and the threshold t in terms of $\boldsymbol{\mu}_j$, σ^2 and γ.

(c) If the covariance matrices are not equal, show that the classifier (8) can be written using a *quadratic discriminant*:

$$\widehat{y} = \begin{cases} 1 & \text{if } g(\mathbf{x}) \geq t, \\ 0 & \text{if } g(\mathbf{x}) < t, \end{cases} \quad g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\intercal \mathbf{Q}\mathbf{x} + \mathbf{c}^\intercal \mathbf{x}. \tag{8}$$

Find $\mathbf{c}$, $\mathbf{Q}$ and t in terms of $\boldsymbol{\mu}_j$, $\mathbf{S}_j$ and γ.