10-702 Statistical Machine Learning: Practice Midterm Exam

Submit solutions to any four of the following seven problems. Clearly indicate which problems you are submitting solutions for. Write your answers in the space provided; additional sheets are attached in case you need extra space.

Problem 1. *Oracles and Consistency*

Let $X \in \mathbb{R}$ and

$$Y = \gamma X^2 + \epsilon \tag{1}$$

where $\mathbb{E}(\epsilon) = 0$.

(a) Find an expression for the oracle linear predictor. In other words, find β_* such that $m(x) = \beta_* x$ minimizes the predictive risk.

(b) We are given n observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (1). Give an estimator $\widehat{\beta}_n$ for β_* and show that it is consistent.

Problem 2. *Model Selection*

Suppose we have the following data:

X	-2	-1	0	1	2
Y	0	0	0	0	0

Consider two regression models:

$$\begin{aligned} &\text{Model 1:} && Y_i = \beta_0 + \epsilon_i \\ &\text{Model 2:} && Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \end{aligned}$$

(a) Use C_p to choose between these model. You may assume that $\sigma^2 = 1$. (Recall that the C_p penalty is $2|S|\sigma^2/n$.)

(b) Fit the model $Y_i = \beta X_i + \epsilon_i$ using the lasso. That is, find $\widehat{\beta}$ to minimize

$$\sum_i (Y_i - \beta X_i)^2 + \lambda|\beta|.$$

Problem 3. *Convex Duality*

Let $X_i \sim \text{Poisson}(\theta)$ be independent, with observations $\{X_1, X_2\} = \{5, 9\}$. Consider the optimization problem

$$\begin{array}{rl} \min\limits_{\theta} & f(\theta) \\ \text{such that} & |\theta| \le 6 \end{array}$$

where $f(\theta) = -\log f_\theta(5) - \log f_\theta(9)$ is the negative log-likelihood, where $f_\theta(n) = \frac{e^{-\theta}\theta^n}{n!}$.

(a) What is the solution to this problem?

(b) Write the Lagrangian.

(c) Derive the dual problem.

(d) State the KKT conditions.

Problem 4. *Convexity and Regularization*

Let Y be the random variable

$$Y = \mu + \epsilon$$

where $\epsilon \sim N(0,1)$ and $\mu \in \mathbb{R}$ is a constant. Define $\widehat{\mu}$ to be the value of μ that minimizes

$$M(\mu) = (Y - \mu)^2 + \lambda J(\mu)$$

where $\lambda > 0$. Consider three cases:

$$\begin{array}{rrcl} 1. & J(\mu) & = & I(\mu \neq 0) \\ 2. & J(\mu) & = & |\mu| \\ 3. & J(\mu) & = & \mu^2. \end{array}$$

(a) For which cases is $M(\mu)$ convex?

(b) Find $\widehat{\mu}$ for all three cases.

Problem 5. *Mixture Models and EM*

Let $(Z_1, X_1, Y_1), \ldots, (Z_n, X_n, Y_n)$ be generated as follows:

$$\begin{array}{rcl} Z_i & \sim & \text{Bernoulli}(p) \\ X_i & \sim & \text{Uniform}(0,1) \\ \epsilon_i & \sim & N(0, \sigma^2) \\ Y_i & \sim & \begin{cases} 5X_i + \epsilon_i & \text{if } Z_i = 0 \\ -5X_i + \epsilon_i & \text{if } Z_i = 1. \end{cases} \end{array}$$

(a) Assume we do not observe the Z_i's or ϵ_i's. Write the distribution $f(x, y)$ of X and Y as a mixture.

(b) Write down the likelihood function for p.

(c) Write down the steps for the EM algorithm.

(d) Find a consistent estimator of p that avoids using EM. Hint: find $\mathbb{E}(Y \mid X = x)$.

Problem 6. *Linear Classification*

Suppose that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}$ and

$$\begin{aligned} X \mid Y = 0 &\sim N(0, 1) \\ X \mid Y = 1 &\sim \frac{1}{2}N(-5, 1) + \frac{1}{2}N(5, 1). \end{aligned}$$

(a) Find an expression for the Bayes classifier and find an expression for the Bayes risk.

(b) What linear classifier minimizes the risk and what is its risk?

Problem 7. *Graphical Models*

Let $X = (X_1, X_2, X_3, X_4, X_5)$ be a random vector distributed as

$$X \sim N(0, \Sigma)$$

where the covariance matrix Σ is given by

$$\Sigma = \frac{1}{15}\begin{pmatrix} 9 & -3 & -3 & -3 & -3 \\ -3 & 6 & 1 & 1 & 1 \\ -3 & 1 & 6 & 1 & 1 \\ -3 & 1 & 1 & 6 & 1 \\ -3 & 1 & 1 & 1 & 6 \end{pmatrix} \quad \text{with inverse} \quad \Sigma^{-1} = \begin{pmatrix} 3 & 1 & 1 & 1 & 1 \\ 1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 0 & 3 \end{pmatrix}$$

(a) What is the graph for X, viewed as a graphical model?

(b) Which of the following independence statements are true?

   (a) $X_2 \perp X_3 \mid X_1$
   (b) $X_3 \perp X_4$
   (c) $X_1 \perp X_3 \mid X_2$
   (d) $X_1 \perp X_5$

(c) List the local Markov properties for this graphical model.

(d) Find the conditional density $p(X_2 \mid X_1 = -3)$.