*Name:* ______________________________

# 10-702 Statistical Machine Learning: Midterm Exam

March 4, 2010

**Submit solutions to any three of the following five 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**. *Convex Duality*

Let $X$ be a $2 \times 2$ symmetric positive definite matrix and let $y_1, \ldots, y_n$ be a collection of vectors in $\mathbb{R}^2$. A problem that arises in statistics and machine learning is to find the minimum volume enclosing ellipsoid centered at the origin of the points $\{y_i\}$. This is expressed as the optimization

$$\begin{array}{rl} \text{minimize} & \log\det X^{-1} \\ \text{subject to} & y_i^T X y_i \leq 1,\ i = 1, \ldots, n. \end{array}$$

(a) Show that this is a convex optimization problem.

(b) Derive the dual function $h(\lambda)$, for $\lambda \in \mathbb{R}^n$.

(c) Give a geometric interpretation for the complementary slackness conditions on $\lambda$.

**Problem 1 work space.**

**Problem 2**. *Graphical Models*

Let $X = (Y, Z) \in \mathbb{R}^6 \sim N(0, \Sigma)$ be a random Gaussian vector where $Y = (Y_1, Y_2) \in \mathbb{R}^2$ and $Z = (Z_1, Z_2, Z_3, Z_4) \in \mathbb{R}^4$, with $\Sigma^{-1} = \Omega = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ where

$$
\begin{aligned}
A &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\
B &= \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ -1 & \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} \end{bmatrix} \\
C &= \begin{bmatrix} 2 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 2 & 0 & 0 \\ 0 & 0 & 2 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 2 \end{bmatrix}.
\end{aligned}
$$

(a) Draw the undirected graph of $X$.

(b) Draw the undirected graph of $Z$. Hint: Recall that

$$
\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}B^TA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}B^TA^{-1} & S^{-1} \end{bmatrix}
$$

where $S = C - B^TA^{-1}B$ is the Schur complement.

(c) Which of the following independence statements hold?

1. $Y_1 \perp Y_2 \,|\, Z$
2. $Z_1 \perp Z_4 \,|\, Z_2$
3. $Z_1 \perp Z_4 \,|\, Y_1$
4. $Z_1 \perp Z_2$

**Problem 2 work space.**

**Problem 3**. *Density Estimation*

Let $X_1, X_2, \ldots, X_T$ be a sequence of identically distributed but *not necessarily independent* random variables, with $X_t \in \mathbb{R}$. Let $\widehat{f}_h(x)$ be the kernel density estimator

$$\widehat{f}_h(x) = \frac{1}{hT} \sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right).$$

Suppose that the covariance of $Z_t \equiv \frac{1}{h} K((X_t - x)/h)$ satisfies

$$\sum_{s \neq t} |\mathrm{Cov}(Z_s, Z_t)| = o(h^{-1})$$

for all $t$. Let $h \to 0$ with $Th \to \infty$.

(a) Derive the asymptotic bias of $\widehat{f}_h(x)$ (to leading order in $h$).

(b) Derive the asymptotic variance of $\widehat{f}_h(x)$.

(c) What is the optimal bandwidth $h$?

State any assumptions you make on the kernel $K$ and the true density $f$.

**Problem 3 work space**.

**Problem 4**. *Regression*

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be $n$ pairs of data where $X_i \in \mathbb{R}$ and $Y_i \in \mathbb{R}$, and denote the true (but unknown) regression function by $m(x) = \mathbb{E}(Y \mid X = x)$.

(a) Consider the class of predictors of the form $\mathcal{G} = \{g(x) = \beta x : \ \beta \in \mathbb{R}\}$. Find the oracle $g_*(x) = \beta_* x$; that is, find $\beta_*$ to minimize the risk $R(\beta) = \mathbb{E}(Y - \beta X)^2$.

(b) Define an estimator $\widehat{g}(x) = \widehat{\beta} x$ and show that $R(\widehat{\beta}) - R(\beta_*) \xrightarrow{P} 0$. You may make any reasonable assumptions on the joint distribution of $(X, Y)$ that you need.

(c) Suppose now that $X_i \in [0, 1]$. Consider the nonparametric estimator

$$\widehat{m}(x) = \sum_{j=1}^{N} \overline{Y}_j \ I(x \in B_j)$$

where $B_1 = [0, 1/N), B_2 = [1/N, 2/N), \ldots, B_N = [(N-1)/N, 1)$, and

$$\overline{Y}_j = \frac{1}{n_j} \sum_{i=1}^{n} Y_i \ I(X_i \in B_j)$$

with $n_j = \sum_{i=1}^{n} I(X_i \in B_j)$. Assume that $n_j > 0$ for all large $n$. Find the approximate bias and variance of $\widehat{m}(x)$ (at a fixed $x$) as a function of $N$. What is the optimal $N$, in terms of mean squared error?

**Problem 4 work space.**

**Problem 5**. *Classification*

Recall that $X \in \mathbb{R}$ has a *nonparanormal* distribution, written $X \sim NPN(\mu, \sigma^2, f)$, in case $f(X) \sim N(\mu, \sigma^2)$. Suppose that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}$ and

$$\begin{aligned} X \,|\, Y = 0 &\sim NPN(0, 1, f) \\ X \,|\, Y = 1 &\sim NPN(1, 1, f). \end{aligned}$$

(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?

(c) Give an algorithm to estimate the best linear classifier from a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$.

**Problem 5 work space.**

*Extra work space.*

*Extra work space.*

*Extra work space.*

*Extra work space.*

*Extra work space.*