Boston University
Department of Electrical and Computer Engineering

ENG EC 500 B1 (Ishwar) Introduction to Learning from Data

Problem Set 2

Issued: Fri 9 Oct 2015 **Due:** 5pm Mon 19 Oct 2015 in box outside PHO440
Required reading: Your notes from lectures and additional notes on website.

Problem 2.1 *(Soft Thresholding)* Let $X = Y + Z$ where $Y \perp\!\!\!\perp Z$, $Z \sim \mathcal{N}(0, \sigma^2)$, $\sigma > 0$, and $Y \sim \pi(y) = 0.5\lambda \exp(-\lambda|y|)$, $\lambda > 0$. Derive $h_{\text{MAP}}(x)$, the MAP estimate of Y based on $X = x$.

Problem 2.2 *(QDA vs LDA)* Consider a binary classification problem in which class labels $y = 0$ and $y = 1$ are equally likely and the feature vector $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$ with the following class-conditional densities:

$$p(\mathbf{x}|y=0) = \begin{cases} \frac{1}{Z_0} & \text{if } ||\mathbf{x}||_1 \leq \frac{1}{\sqrt{2}} \\ 0 & \text{otherwise.} \end{cases} \qquad p(\mathbf{x}|y=1) = \begin{cases} \frac{1}{Z_1} & \text{if } ||\mathbf{x} - (\sqrt{2}, 0)^\top||_1 \leq \sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the values of Z_0 and Z_1.

(b) Compute the MPE rule (Bayes rule for the 0-1 loss function).

(c) Compute the Bayes risk (misclassification probability) corresponding to the Bayes rule from part (b).

(d) Compute the class-conditional mean vectors $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$ and covariance matrices Σ_0, Σ_1 of the feature vectors.

(e) Even though not Gaussian, suppose that we model the class-conditional pdfs as Gaussian with mean vectors and covariance matrices respectively given by $\boldsymbol{\mu}_y$, Σ_y for $y = 0, 1$. (i) Compute the QDA decision rule and simplify it as much as possible. (ii) Compute the resulting misclassification probability.

(f) Now suppose that we use the LDA model instead where the class-conditional pdfs are Gaussian with mean vectors given by $\boldsymbol{\mu}_y$, $y = 0, 1$ and a *common* covariance matrix given by $\Sigma = P(Y = 0)\Sigma_0 + P(Y = 1)\Sigma_1$. (i) Compute the LDA decision rule and simplify it as much as possible. (ii) Compute the resulting misclassification probability.

(g) How does the decision rule and misclassification probability change compared to part (e) if we used a Gaussian Naive Bayes model instead?

Problem 2.3 *(Comparing k-NN performance)* Let $S := \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}||_2 \leq 1\}$ be a spherical ball of radius 1 in $\mathbb{R}^d$. Consider a binary classification problem in which the class labels are equally likely, i.e., $P(Y = 0) = P(Y = 1) = 0.5$, and the feature vector $x \in \mathbb{R}^d$ with the following class-conditional densities:

$$p(x|y) = \text{Uniform}(\boldsymbol{\mu}_y + S), \; y = 0, 1,$$

with $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\| > 4$. Let $\mathcal{D} := \{(\mathbf{X}_j, Y_j), j = 1, \ldots, n\}$ be n IID labeled training examples with joint distribution $1/2\,p(\mathbf{x}|y)$. Let $(\mathbf{X}_{\text{test}}, Y_{\text{test}})$ be a test pair which is drawn independently of $\mathcal{D}$ according to the same joint distribution $0.5p(\mathbf{x}|y)$. Let $h_{k-\text{NN}}(\mathbf{x}; \mathcal{D})$ denote the k-NN decision rule based on $\mathcal{D}$, where k is an odd positive integer. Since $\mathcal{D}$ is random, the decision rule $h_{k-\text{NN}}(\mathbf{x}; \mathcal{D})$ evaluated at any point $\mathbf{x}$ is a random variable.

(a) Compute the MAP rule and its misclassification probability.

(b) Compute $P(h_{k-\text{NN}}(\mathbf{X}_{\text{test}}; \mathcal{D}) \neq Y_{\text{test}})$ for $k = 1, 3, 5, \ldots$.

(c) Compare and order the performance of the k-NN rule for $k = 1, 3, 5, \ldots$ from best to worst.

(d) Evaluate the misclassification probability of the k-NN rule for each k as $n \to \infty$.