

ECE 592 – Topics in Data Science

Test 4: Machine Learning – Fall 2020

October 28, 2020

Please remember to justify your answers carefully.

Last name: _____ First name: _____

Question 1 (Bayesian classification)

Consider a Bayesian classification problem where there are two classes, red and blue. The probabilities of the two classes are $Pr(\text{blue}) = 0.6$ and $Pr(\text{red}) = 0.4$. A random variable (RV) X is generated in different ways based on the class. For each class, the conditional probability density function (pdf) is a Gaussian mixture with 2 components, $f_{\text{blue}} = 0.5\mathcal{N}(0, 1) + 0.5\mathcal{N}(1, 1)$ and $f_{\text{red}} = 0.8\mathcal{N}(0, 1) + 0.2\mathcal{N}(1, 1)$, where the means of all Gaussian components are 0 and 1, the variances corresponding to all Gaussian components are 1, and we have probabilities 0.2, 0.5, and 0.8 for the components. Recall that a Gaussian RV X with mean μ and variance σ^2 has a pdf given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

(a) Derive a Bayesian classifier by computing the posterior probability, $Pr(\text{red}|x)$.

Solution:

$$\begin{aligned} Pr(\text{red}|x) &= \frac{f(\text{red}, x)}{f(x)} \\ &= \frac{Pr(\text{red})f(x|\text{red})}{Pr(\text{red})f(x|\text{red}) + Pr(\text{blue})f(x|\text{blue})} \\ &= \frac{0.4(0.8 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0.2 \frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}})}{0.4(0.8 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0.2 \frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}}) + 0.6(0.5 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0.5 \frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}})}. \end{aligned}$$

(b) Find the decision boundary/boundaries for the classifier that you derived above? (There is no need to provide numerical values; an expression is fine.)

Solution: For the decision boundary, we need to compute the value of x for which $Pr(\text{red}|x) = 0.5$. For the probability to be equal, the two expressions in the denominator in the solution

of part (a) must be identical,

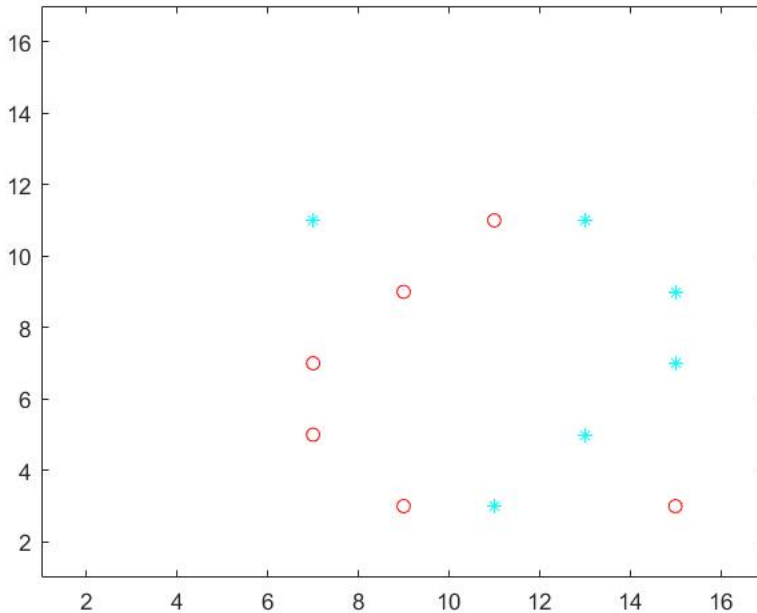
$$\begin{aligned}0.4\left(0.8\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0.2\frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}}\right) &= 0.6\left(0.5\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0.5\frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}}\right) \\(0.4 \cdot 0.8 - 0.6 \cdot 0.5)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} &= (0.6 \cdot 0.5 - 0.4 \cdot 0.2)\frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}} \\0.02\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} &= 0.22\frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}} \\e^{-\frac{x^2}{2}} &= 11\frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}} \\e^{-\frac{x^2}{2}} &= 11e^{-\frac{(x-1)^2}{2}} \\-\frac{x^2}{2} &= \ln(11) - \frac{(x-1)^2}{2}.\end{aligned}$$

Question 2 (Nearest neighbors classification)

You are given training vectors labeled into two classes, where each vector is in $p = 2$ dimensions. The training set is comprised of the following points, which are plotted below.

Class 1 (red circles): [(11, 11); (15, 3); (9, 9); (7, 7); (7, 5); (9, 3)]

Class 2 (blue asterisks): [(15, 9); (15, 7); (13, 5); (13, 11); (11, 3); (7, 11)]



Please answer the following questions.

(a) For the K nearest neighbors classifier, what value of K seems reasonable for this dataset? What is the resulting training error? (Because there might be different reasonable answers, make sure to justify your response. Note that the training error in this question is the fraction of mis-classified vectors in the training data.)

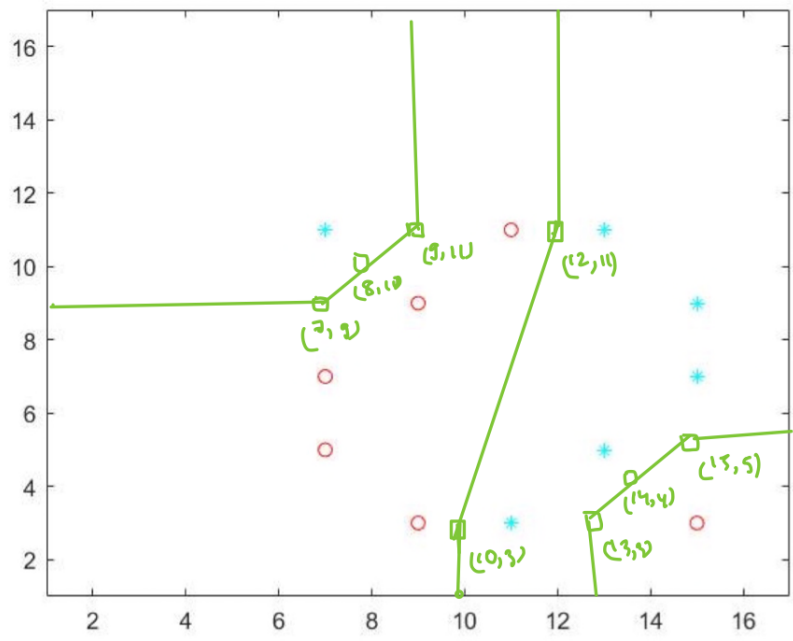
Solution:

- If the points on the top left and bottom right are considered outliers, then $3 \leq K \leq 6$ will yield the least error. For $K = 3$, the training error will be $\text{err} = 2/12 = 0.16$.
- If points are not considered outliers, then any small value of K will give zero training error.

(b) Are there some values of K that might be too large or too small for this dataset?

Solution: Small values, e.g., $K = 1$, may cause over-fitting. Similarly, large values of k may cause under-fitting, resulting in incorrect classification.

(c) Please sketch the 1-nearest neighbor decision boundary for this dataset.



Question 3 (LASSO)

Consider a vector β observed through noisy measurements y ,

$$y = X\beta + z. \quad (1)$$

Our goal is to recover or estimate $\beta \in \mathbb{R}^N$, given $X \in \mathbb{R}^{M \times N}$ and $y \in \mathbb{R}^M$. Below, you will show in several steps that when β and z are modeled as independent and identically distributed (i.i.d.) Gaussian, maximum a posteriori (MAP) estimation of β ,

$$\beta_{MAP} = \arg \max_{\beta} f(\beta|X, y),$$

is a special case of the least absolute shrinkage and selection operator (LASSO). (Note that this problem is closely related to a problem from the 2019 final exam, and some parts that students had to explain in that exam are not required on this one.)

We begin deriving the solution β_{MAP} that maximizes $f(\beta|X, y)$,

$$\beta_{MAP} = \arg \max_{\beta} f(\beta|X, y) = \arg \max_{\beta} \frac{f(\beta, X, y)}{f(X, y)} = \arg \max_{\beta} f(\beta, X, y).$$

Focusing on the last term, $f(\beta, X, y) = f(X)f(\beta|X)f(y|\beta, X)$, but β is independent of X , i.e., $f(\beta|X) = f(\beta)$, and so

$$\beta_{MAP} = \arg \max_{\beta} f(\beta, X, y) = \arg \max_{\beta} f(X)f(\beta)f(y|\beta, X) = \arg \max_{\beta} f(\beta)f(y|\beta, X). \quad (2)$$

(a) To compute $f(\beta)$ and $f(y|\beta, X)$ in (2), we model each of the M scalar entries in $z \in \mathbb{R}^M$ (1), which can be interpreted as a noise or error vector, as i.i.d. zero-mean Gaussian with variance σ_Z^2 . That is, $Z_m \sim \mathcal{N}(0, \sigma_Z^2)$, where Z_m is the random variable corresponding to entry m of the noise vector, and $m \in \{1, \dots, M\}$. The pdf for Z_m can be expressed,

$$f(Z_m = z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{z^2}{2\sigma_Z^2}}.$$

You are also given that the N entries of β are i.i.d. with pdf

$$f(\beta_n) = c_1 e^{c_2 |\beta_n|},$$

where $n \in \{1, \dots, N\}$, and $c_1 > 0$ and $c_2 < 0$ are constants that control the variance of this RV. Derive expressions for $f(\beta)$ and $f(y|\beta, X)$ in terms of X, y, σ_z, c_1, c_2 . (Hint: you can simplify your expressions using ℓ_1 and ℓ_2 norm notations for β and $y - X\beta$.)

Solution:

$$f(\beta) = \prod_{n=1}^N f(\beta_n) = \prod_{n=1}^N [c_1 e^{c_2 |\beta_n|}] = (c_1)^N \exp \left\{ c_2 \sum_{n=1}^N |\beta_n| \right\} = (c_1)^N \exp \{c_2 \|\beta\|_1\}.$$

$$\begin{aligned} f(y|\beta, X) &= f(Z = y - X\beta) \\ &= \prod_{m=1}^M \left[\frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp \left\{ -\frac{([y - X\beta]_m)^2}{2\sigma_Z^2} \right\} \right] \\ &= (2\pi\sigma_Z^2)^{-M/2} \exp \left\{ -\frac{\|y - X\beta\|_2^2}{2\sigma_Z^2} \right\}. \end{aligned}$$

(b) Because the logarithm is a monotone function, it suffices to maximize $\log(f(\beta|y, X))$,

$$\begin{aligned}\beta_{MAP} &= \arg \max_{\beta} f(\beta)f(y|\beta, X) \\ &= \arg \max_{\beta} \log(f(\beta)f(y|\beta, X)) \\ &= \arg \max_{\beta} \text{Polynomial}(\beta, y, X, \sigma_Z, c_1, c_2).\end{aligned}$$

Express $\text{Polynomial}(\beta, y, X, \sigma_Z, \sigma)$.

Solution: Substituting the solutions from part (a),

$$f(\beta)f(y|\beta, X) = (c_1)^N \exp\{c_2\|\beta\|_1\} (2\pi\sigma_Z^2)^{-M/2} \exp\left\{-\frac{\|y - X\beta\|_2^2}{2\sigma_Z^2}\right\}.$$

The log satisfies,

$$\log(f(\beta)f(y|\beta, X)) = \text{const} + c_2\|\beta\|_1 - \frac{\|y - X\beta\|_2^2}{2\sigma_Z^2}.$$

(c) The LASSO has the form

$$\beta_{\text{LASSO}} = \arg \min_{\beta} \|y - X\beta\|^2 + \lambda\|\beta\|_1,$$

where $\|\cdot\|^2$ is the squared ℓ_2 norm, and $\|\cdot\|_1$ is the ℓ_1 norm. The LASSO form should correspond to your expression above; what is λ ? What intuition can you draw from the expression for λ ? If you could not derive the expression for β_{MAP} , please explain how λ impacts the LASSO solution.

Solution: First, note that $c_2 < 0$, hence there is no problem with the sign. Next, multiplying the polynomial of part (b) by $-2\sigma_Z^2$,

$$\widetilde{\text{polynomial}} = \widetilde{\text{const}} - 2c_2\sigma_Z^2\|\beta\|_1 + \|y - X\beta\|_2^2.$$

Therefore, $\lambda = -2c_2\sigma_Z^2$.