Georgia Institute of Technology

ECE 6254 - Midterm

Fall 2020 - v1.0

- There are 4 problems over 5 pages (including the cover page)
- The problems are not necessarily in order of difficulty.
- All problems are assigned the same weight in the overall grade. The questions in problems with many questions are worth less than the questions in problem with few questions.
- Each question is graded as follows: no credit without meaningful work, half credit for partial work, full credit if essentially correct.
- Unless otherwise specified, you should concisely indicate your reasoning and show all relevant work.
- The grade on each problem is based on our judgment of your level of understanding as reflected by what you have written. If we can't read it, we can't grade it.
- This is an open-note exams. All materials posted on canvas are fair game, but use of other resources is not.
- I'm trusting you to behave honorably and abide by the spirit of the honor code. Evidence of cheating will be handled with extreme prejudice, most likely by cancelling the midterm and replacing it by an impossibly hard exam.
- Please use a pen and not a pencil if you handwrite your solution.
- By starting this exam, you agree with the terms and conditions listed above
- **You must submit your exam on Gradescope. Make sure you allocate time for the submission**.

Problem 1: Bayes classifier with Laplace distributions

Consider a binary classification problem involving a single (scalar) feature x and suppose that $X|Y = 0$ and $X|Y = 1$ are continuous random variables with densities given by

$$g_0(x) = \frac{1}{2} e^{-|x|}$$
$$g_1(x) = \frac{1}{2} e^{-|x-1|},$$

respectively. Furthermore, suppose that $\mathbb{P}[Y = 0] = \mathbb{P}[Y = 1] = \frac{1}{2}$.

[Q1] Derive the optimal classification rule (in terms of minimizing the probability of error)

[Q2] Calculate the Bayes risk for this classification problem when using the 0/1 loss.

Problem 2: Bayes Optimal Classifiers

In classification, the loss function we usually want to minimize is the regular 0/1 loss: $\ell(f(x), y) = \mathbb{1}\{f(x) \neq y\}$ where $f(x), y \in \{0, 1\}$ (i.e., binary classification). In this problem we will consider the effect of using an asymmetric loss function:

$$\ell_{\alpha,\beta}(f(x), y) = \alpha \mathbb{1}\{f(x) = 1, y = 0\} + \beta \mathbb{1}\{f(x) = 0, y = 1\}.$$

Under this loss function, the two types of error receive different weights, determined by $\alpha, \beta > 0$.

[Q1] Determine the Bayes optimal classifier, i.e. the classifier that achieves minimum risk assuming $P(x, y)$ is known, for the loss $\ell_{\alpha,\beta}$ where $\alpha, \beta > 0$.

[Q2] Suppose that the class $y = 0$ is extremely uncommon (i.e., $P(y = 0)$ is small). This means that the classifier $f(x) = 1$ for all x will have good risk. We may try to put the two classes on even footing by considering the risk:

$$R(f) = P(f(x) = 1|y = 0) + P(f(x) = 0|y = 1).$$

Show how minimizing $R(f)$ is equivalent to choosing a certain α, β and minimizing $\mathbb{E}(\ell_{\alpha,\beta}(f(x), y))$ for this specific α, β.

Problem 3: Playing with classifiers

The objective of this problem is to analyze a synthetic dataset and visualize the result of classification in python. You are allowed to use the functions provided by the libraries `scikit-learn`, `numpy`, but use of other libraries is not allowed. In every question where code is required, *include* the code in your answer. In every question where a graph is required, *provide* a brief discussion. Answers without discussion will not receive full credit.

[Q1] Generate a training dataset consisting of $N = 100$ feature vectors and binary label pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$

according to the following distribution.

$$\mathbb{P}(y = 0) = \frac{1}{3} \quad \mathbb{P}(y = 1) = \frac{2}{3} \tag{1}$$

$$p(\mathbf{x}|y = 0) \sim \mathcal{N}\left(\mu_0 \triangleq [0, 1]^\intercal, \Sigma_0 \triangleq \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}\right) \tag{2}$$

$$p(\mathbf{x}|y = 1) \sim \mathcal{N}\left(\mu_1 \triangleq [0, -2]^\intercal, \Sigma_1 \triangleq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{3}$$

[Q2] Generate a testing dataset consisting of N_{test} feature vectors and binary label pairs. Our objective will be to use the testing set to evaluate the risk of classifiers. Explain how you will choose the value of N_{test} to obtain an accurate estimate of the true risk. This is a conceptual question but your are expected to ground your answer in some of the mathematical results we've used in class.

[Q3] Using the testing dataset, write a function to evaluate the risk (with a 0/1 loss) of a generic binary classifier.

[Q4] Write an expression for the Bayes classifier using the information in part **[Q1]**. *Simplify your result as much as possible.*

[Q5] Implement the Bayes classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier on the testing set.

[Q6] Implement a K nearest neighbor classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier, explain how you chose the value of K.

[Q7] Implement an LDA classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier, explain the results that you observe.

[Q8] A QDA classifier is similar to an LDA classifier, but the covariance matrices of the generative distributions are now allowed to depend on the class label. Implement a QDA classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier, explain the results that you observe.

[Q9] Implement a logistic regression classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier, explain the results that you observe.

[Q10] Implement a Gaussian Naive Bayes classifier and show its behavior on a pretty graph. At the very least, your result should clearly show the decision boundary, the decision regions of the classifier, and the original training set. Report the risk of the classifier, explain the results that you observe.

Problem 4: Yes/No questions

(1) For a continuous random variable x and its probability distribution function $p(x)$, it holds that $0 \leqslant p(x) \leqslant 1$ for all x

☐ Yes ☐ No

Justification:

(2) The training error of 1-NN classifier is 0.
☐ Yes ☐ No
Justification:

(3) No classifier can do better than a naive Bayes classifier if the distribution of the data is known.
☐ Yes ☐ No
Justification:

(4) Maximizing the likelihood of logistic regression model yields multiple local optima.
☐ Yes ☐ No
Justification:

(5) PAC learnability requires one to know the distribution that generates the data
☐ Yes ☐ No
Justification:

(6) The kernel trick allows on to operate in a different Hilbert space by computing transformed feature vectors
☐ Yes ☐ No
Justification:

(7) $p(x, y|z) = p(x|z)p(y|z)$ if $p(x|y, z) = p(x|z)$
☐ Yes ☐ No
Justification:

(8) When the hypothesis space is richer, overfitting is more likely.
☐ Yes ☐ No
Justification:

(9) Gaussian Naive Bayes for two classes will always give a linear decision boundary.
☐ Yes ☐ No
Justification:

(10) Logistic regression for two classes will always give a linear decision boundary.
☐ Yes ☐ No
Justification: