Engineering Applications of Machine Learning and Data Analytics
Homework #1 [Due: 01/28/2022]

I acknowledge that this homework assignment is solely my effort. I have done this work by myself. I have not consulted with others about this exam in any way. I have not received outside aid (outside of my own brain) on this exam. I understand that violation of these rules contradicts the class policy on academic integrity.

Name: _____

Signature: _____

Date: _____

Instructions: There are seven problems. Partial credit is given for answers that are partially correct. No credit is given for answers that are wrong or illegible. Write neatly. You must submit two PDFs on D2L. The first PDF has the results to the analytical questions as well as figures that are generated

Problem 1: _____

Problem 2: _____

Problem 3: _____

Total: _____

1   Probability and Discriminant Classifiers [20pts]

PART I: Maximum Posterior vs Probability of Chance   Show/explain that $P(\omega_{\max}|\mathbf{x}) \geq \frac{1}{c}$ when we are using the Bayes decision rule, where c is the number of classes. Derive an expression for $p(\text{err})$. Let $\omega_{\max}$ be the state of nature for which $P(\omega_{\max}|\mathbf{x}) \geq P(\omega_i|\mathbf{x})$ for $i =, 1 \ldots, c$. Show that $p(\text{err}) \leq (c-1)/c$ when we use the Bayes rule to make a decision. Hint, use the results from the previous questions.

PART II: Bayes Decision Rule Classifier   Let the elements of a vector $\mathbf{x} = [x_1, \ldots, x_d]^\mathsf{T}$ be binary valued. Let $P(\omega_j)$ be the prior probability of the class ω_j ($j \in [c]$), and let

$$p_{ij} = P(x_i = 1|\omega_j)$$

with all elements in $\mathbf{x}$ being independent. If $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, and $p_{i1} = p > \frac{1}{2}$ and $p_{i2} = 1 - p$, show that the minimum error decision rule is

$$\text{Choose } \omega_1 \text{ if } \sum_{i=1}^{d} x_i > \frac{d}{2}$$

Hint: Think back to ECE503 and types of random variables then start out with

$$\text{Choose } \omega_1 \text{ if } P(\omega_1)P(\mathbf{x}|\omega_1) > P(\omega_2)P(\mathbf{x}|\omega_2)$$

PART III: The Ditzler Household Growing Up   My parents have two kids now grown into adults. Obviously there is me, Greg. I was born on a Wednesday. What is the probability that I have a brother? You can assume that $\mathbb{P}(\text{boy}) = \mathbb{P}(\text{girl}) = \frac{1}{2}$.

PART IV: Bayes classifier   Let consider a Bayes classifier with $p(\mathbf{x}|\omega_i)$ distributed as a multivariate Gaussian with mean μ_i and covariance $\Sigma_i = \sigma^2 I$ (note they all share the same covariance). We choose the class that has the largest

$$g_i(\mathbf{x}) = \log(p(\mathbf{x}|\omega_i)P(\omega_i)) \propto \mathbf{w}_i^\mathsf{T}\mathbf{x} + w_{0i}$$

Find $\mathbf{w}_i$ and w_{0i}. Fact:

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_i)^\mathsf{T}\Sigma_i^{-1}(\mathbf{x} - \mu_i)\right\}$$

2 Linear and Quadratic Classifiers – Code [20pts]

- Write a general function to generate random samples from $\mathcal{N}(\mu, \Sigma)$ in d-dimensions (i.e., $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$).

- Write a procedure of the discriminant of the following form

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^{\mathsf{T}} \Sigma_i^{-1}(\mathbf{x} - \mu_i) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i)) \qquad (1)$$

- Generate a 2D dataset with three classes and use the quadratic classifier above to learn the parameters and make predictions. As an example, you should generate training data shown below to estimate the the parameters of the classifier in (1) and you should test the classifier on another randomly generated dataset. It is also sufficient to show the dataset used to train your classifier and the decision boundary it produces.
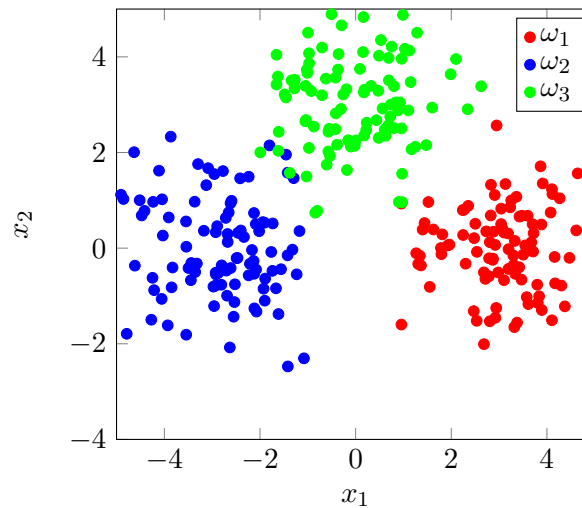
- Write a procedure for computing the Mahalanobis distance between a point $\mathbf{x}$ and some mean vector μ, given a covariance matrix Σ.

- Implement the naïve Bayes classifier from scratch and then compare your results to that of Python's built-in implementation. Use different means, covariance matrices, prior probabilities (indicated by relative data size for each class) to demonstrate that your implementations are correct.

3   Misc Code [20pts] - Choose One

Problem I: Comparing Classifiers   A text file, `hw1-scores.txt`, containing classifier errors measurements has been uploaded to D2L. Each of the columns represents a classifier and each row a data set that was evaluated. Are all of the classifiers performing equally? Is there one or more classifiers that is performing better than the others? Your response should be backed by statistics. Suggested reading:

- Janez Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," Journal of Machine Learning Research, vol. 7, 1–20.

Read the abstract to get an idea about the theme of comparisons. Sections 3.1.3 and 3.2.2 can be used to answer the question posed here.

Problem II: Sampling from a Distribution   Let the set $\mathcal{N} \in [1, \ldots, n, \ldots, N]$ be a set of integers and $\mathbf{p}$ be a probability distribution $\mathbf{p} = [p_1, \ldots, p_n, \ldots, p_N]$ such that p_k is the probability of observing $k \in \mathcal{N}$. Note that since $\mathbf{p}$ is a distribution then $\mathbf{1}^\mathsf{T}\mathbf{p} = 1$ and $0 \leq p_k \leq 1 \; \forall n$. Write a function $\mathtt{sample}(M, \mathbf{p})$ that returns M indices sampled from the distribution $\mathbf{p}$. Provide evidence that your function is working as desired. Note that all sampling is assumed to be i.i.d. You must include a couple of paragraphs and documented code that discusses how you were able to accomplish this task.