ECE523: Engineering Applications of Machine Learning and Data Analytics

I acknowledge that this exam 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 five problems. You have 50 minutes to complete the exam. Partial credit is given for answers that are partially correct. No credit is given for answers that are wrong or illegible. Write neatly.

Problem 1: __________

Problem 2: __________

Problem 3: __________

Problem 4: __________

Problem 5: __________

Total: __________

Problem #1 – Ridge Regression (10 Points)

In class we discussed linear discriminant models and one approach was linear regression. In this problem we look at ridge regression, which is given by

$$\arg\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2 + \frac{\lambda}{2}\|\mathbf{w}\|_2^2$$

where $\mathbf{y} \in \mathbb{R}^n$ is a vector of the outputs, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the matrix of data and $\mathbf{w} \in \mathbb{R}^p$ are the parameters for the linear model $y = \mathbf{w}^\mathsf{T}\mathbf{x}$. Find $\mathbf{w}$.

Problem #2 – Principal Component Analysis (10 Points)

In class, we showed two different approaches that we could arrive at a solution to PCA: one with linear algebra and one with optimization. This problem asks you to use both of what you know about the PCA projection and task of optimization. Use these facts:

- The projection is performed with $z = \mathbf{w}^\mathsf{T}\mathbf{x}$. Note that z is a scalar because we are only looking for one principal axis.
- I am not too concerned with the magnitude of $\mathbf{w}$, but I am concerned with its direction.
- You need to maximize the variance of z.

Use these facts to find $\mathbf{w}$. It maybe a good idea to let $\mathbf{X} \in \mathbb{R}^{p \times n}$ be the matrix of data. Then the covariance matrix is given by $\frac{1}{n-1}\mathbf{X}\mathbf{X}^\mathsf{T} = \Sigma$. This approach is similar to how we discussed PCA from a linear algebra perspective.

Problem #3 – A Gamblers Ruin (10 Points)

[True/False] (1 point): Density estimation (using say, the kernel density estimator) can be used to perform classification.

[True/False] (1 point): One of the disadvantages of the logistic function is that its derivative is not very convenient to compute.

[True/False] (1 point): Logistic regression assumes that the log-likelihood ratio for two classes with equal priors is linear. More formally this is given by

$$\log\left\{\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}\right\} = \mathbf{w}^\mathsf{T}\mathbf{x} + w_0$$

[True/False] (1 point): Regularization is one way to prevent overfitting and the reason it is so effective is because the regularization term is data-dependent. Therefore, the optimization process will "find" the best way to be resilient against overfitting.

[True/False] (1 point): The training error of 1-NN classifier is 0.

[True/False] (1 point): The principal components are the ones that maximize the variance within a class.

[True/False] (1 point): The correspondence between logistic regression and naïve Bayes (with identity class covariances) means that there is a one-to-one correspondence between the parameters of the two classifiers.

[True/False] (1 point): The number of actions need not be equal to the number of classes. This question is in the context of risk and decision making with Bayes.

[True/False] (1 point): I don't like true and false questions, but I do like free points!

[Accept/Reject] (1 point): "My algorithm is better than yours. Look at the training error rates!"

[Accept/Reject] (1 point): "My algorithm is better than yours. Look at the training error rates and the p-value from the signed rank Wilcoxon test! (Footnote: reported results for best value of λ, chosen with 10-fold cross validation.)"

Problem #4 – To Bayes or Not Bayes (10 Points)

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:

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

Hints: Start with $g_i(\mathbf{x})$ and the fact stated above. Then begin to drop out the terms that are constant for all $g_i(\mathbf{x})$ to simplify the solution.

Problem #5 – Density Estimation (10 Points)

In class, we discussed three conditions that need be met if a density estimator ($p_n(\mathbf{x}) = \frac{k_n/n}{V_n}$) is to converge in probability to the true density ($p(\mathbf{x})$). More formally,

$$\lim_{n\to\infty} V_n = 0, \quad \lim_{n\to\infty} k_n = \infty, \quad \lim_{n\to\infty} \frac{k_n}{n} = 0$$

where k_n is the number of samples that fall within a region $\mathcal{R}$ with volume V_n. Describe two out of the three conditions and why they are necessary for $p_n(\mathbf{x})$ to converge in probability to $p(\mathbf{x})$ when n approaches inifinity.