Math 156 Term Project 2

Deadline: Tuesday May 5th 11:59pm

Problem 1

Recall in linear Bayesian regression, we assume the data set $\mathcal{D} = \{(\mathbf{x}_i, t_i)\}_{i=1}^N$ satisfies a pattern

$$t_i = f(\mathbf{x}_i) + \epsilon_i$$

for some function $f : \mathbb{R}^D \to \mathbb{R}$. We assume the noises $\{\epsilon_i\}_{i=1}^N$ are i.i.d. $\sim \mathcal{N}(0, \beta^{-1})$. We wish to approximate the function $f(\mathbf{x})$ with a function $y(\mathbf{x}, \mathbf{w})$,

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}),$$

where $\mathbf{x} \in \mathbb{R}^M$ is a coefficient vector, and $\boldsymbol{\phi}(\mathbf{x}) = (1, \mathbf{x})$ (the *basis function*). We then view $\mathbf{w}$ as a random variable and assume the distribution $\mathbf{w} \sim N(\mathbf{0}, \alpha^{-1}\mathbf{I})$ (the *prior*). By Bayes' Theorem, the probability of the coefficients $\mathbf{w}$ given the training data set $\mathcal{D} = \{(\mathbf{x}_i, t_i)\}_{i=1}^N$ (the *posterior probability*) is

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}. \tag{1}$$

To approximate $f(\mathbf{x})$ with $y(\mathbf{x}, \mathbf{w})$, we search for the optimal coefficient $\mathbf{w}^*$ that maximizes the posterior probability.

(a) Explain why we can search for $\mathbf{w}^*$ that maximizes $p(\mathcal{D}|\mathbf{w})p(\mathbf{w})$ instead of the expression in Equation (1). (5 points)

(b) The *likelihood function*, as a result of the noise distribution, is given by

$$p(\mathcal{D}|\mathbf{w}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp\left(\frac{|t_i - y(\mathbf{x}_i, \mathbf{w})|^2}{2\beta^{-1}}\right).$$

Combine this and the distribution of $\mathbf{w}$, find the conditional expectation $\mathbb{E}[\mathbf{w}|\mathcal{D}]$ and conditional covariance matrix $\text{Var}[\mathbf{w}|\mathcal{D}]$. (15 points)
Hint: Multiplying two Gaussian $p(\mathcal{D}|\mathbf{w})$ and $p(\mathbf{w})$ results to an Gaussian p.d.f.-like function, hence the posterior $p(\mathbf{w}|\mathcal{D})$ is also Gaussian. Try to write $p(\mathcal{D}|\mathbf{w})p(\mathbf{w})$ into the form of a Gaussian p.d.f. and complete the squares to figure out the mean and covariance matrix.

Problem 2

In Section 4.1, we explore classification with linear discriminants, that is, the decision boundary are hypersurfaces. For a K-class classification problem, we construct K linear functions

$$y_{[k]}(\mathbf{x}) = \mathbf{w}_{[k]}^T \mathbf{x} + w_{0,[k]}$$

where the weights $\mathbf{w}_{[k]}$ and biases $w_{0,[k]}$ (for each class $k = 1, \cdots, K$) are to be determined. The decision regions $\mathcal{R}_{[k]}$ are given by

$$\mathcal{R}_{[k]} = \left\{ \mathbf{x} \in \mathbb{R}^D \mid \arg\max_{\kappa} y_{[\kappa]}(\mathbf{x}) = k \right\}.$$

(Note that κ and k are two different symbols though visually very similar.) Prove that these decision regions $\mathcal{R}_{[k]}$ are convex. (10 points)

Problem 3

Visit the **UCI Machine Learning Repository**. The UCI Machine Learning Repository is a collection of databases (and more) that are used by the machine learning community for the empirical analysis of machine learning algorithms. It is strongly suggested that you find inspiration for your final project from this repository. (Be mindful of their citation policy.) Look for data sets with Classification as their Default Task. Select a data set with #instances ≤ 1000 (this is to ensure the task doesn't get harder than the scope of term projects) and make sure you **cite the data set at the end of your submission**. I suggest you use the Breast Tissue Data Set unless you're specifically motivated for any particular data set.

(a) Write down the mathematical plan for your regression/classification on the data set. You can use the basis function $\boldsymbol{\phi}(\mathbf{x}) = (1, \mathbf{x})$ from linear regression and an estimator $y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$. What is your minimization problem? Provide a statistical model behind it. (30 points)

(b) Download and read the data into your program. (10 points)

(c) Solve the problem numerically. That is, solve the optimization problem and find the optimal parameter $\mathbf{w}^*$. (15 points) Use the firsts 50% of data for training and the rest 50% for validation. Verify the your result using the test data and report your accuracy rate. You can use any existing library to perform this task as long as you can explain its mathematical structure in task (a). (15 points)

(d) **(Bonus)** Use different strategies to divide the data set into training and validation sets. How does your accuracy rate change? (up to 15 points)

Important submission note: (same as term project # 1) You can use any programming language to accomplish the coding tasks. The university now provides MATLAB

access to all UCLA students, see https://www.it.ucla.edu/news/matlab-software-now-available. Other high-level programming languages suitable for our tasks include Python, R, Julia, Octave, and more. The instructor and the TA guarantee assistance with MATLAB programming and can potentially help you trouble shoot your code in other programming languages (not guaranteed). If you wish to email us asking for programming assistance, please attach your code, highlight the part in question, and articulate your problems. Please attach your code at the end of your pdf submission as plain text and document a list of your collaborators and external resource. This document is created with LaTeX and you can find the source .tex file on CCLE. For more LaTeX help, check out this LaTeX tutorial link, or take a look at my template .tex file.