

Homework #1

CS 260: Machine Learning Algorithms

Prof. Ameet Talwalkar

Due: 10/6/15, 8am

1 Sequence of Coin Flips

Suppose you have a *biased* coin with probability of heads equal to p . Imagine that you flip this coin until observing the first heads. Let X denote the number of flips needed to observe the first heads.

- Describe $\mathbb{P}[X = k]$ as a function of p .
- For any $x_0 \geq 1$, show that the probability $\mathbb{P}[X \geq x_0] = (1 - p)^{x_0 - 1}$.
- Suppose we have a prior belief that p is uniformly distributed in the $[0, 1]$ interval. Assuming that the first coin flip equals heads, compute the probability that $p > 1/2$, i.e., compute $\mathbb{P}[p > 1/2 | X = 1]$. Does our belief about the probability of the event $\{p > 1/2\}$ increase or decrease after observing the head in the first flip. Hint: Use Bayes Rule.

2 Convex Functions and Information Theory

- Show that the function $f(x) = |x| + \exp(x)$ is convex.
- Suppose the random variable X is distributed according to a k -class multi-nominal distributions with class probabilities p_1, p_2, \dots, p_k , such that $\sum_{i=1}^k p_i = 1$. Find the values of $p_i, i = 1, \dots, k$ such that the **entropy** of X is maximized.

3 Linear Algebra

- The covariance matrix Σ of a random vector \mathbf{X} is defined as $\Sigma = \mathbb{E}[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^T]$, where $\mathbb{E}\mathbf{X}$ is the expectation of \mathbf{X} . Is Σ **positive-semidefinite**?
- Let \mathbf{A} and \mathbf{B} be two $\mathbb{R}^{D \times D}$ symmetric matrices. Suppose \mathbf{A} and \mathbf{B} have the exact same set of **eigenvectors** $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_D$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_D$ for \mathbf{A} , and $\beta_1, \beta_2, \dots, \beta_D$ for \mathbf{B} . Please write down the eigenvectors and their corresponding eigenvalues for the following matrices:
 - $\mathbf{C} = \mathbf{A} + \mathbf{B}$
 - $\mathbf{D} = \mathbf{A} - \mathbf{B}$
 - $\mathbf{E} = \mathbf{AB}$
 - $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ (assume \mathbf{A} is invertible)

4 KNN Classification in MATLAB/Octave

In this problem, you will implement a KNN classifier and deploy it on a real-world dataset. Below, we describe the steps that you need to take to accomplish this programming assignment.

You will work with a preprocessed version of the *Car Evaluation Dataset* from UCI's machine learning data repository. The training/validation/test sets are provided along with the assignment as `cars_train.data`, `cars_valid.data`, and `cars_test.data`. For a description of the dataset and to determine which field corresponds to the label, please refer to <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>.

- The first step in every data analysis experiment involves inspecting the data and to make sure it is properly formatted. You will find that the features in the provided dataset are categorical. However, KNN requires the features to be real-valued numbers. To convert a categorical feature with K categories to a real-valued number, you can create K new *binary* features. The i th binary feature indicates whether the original feature belongs to the i th category or not. This strategy is called 'one-hot encoding.'
- Please fill in the function `knn_classify` in `knn_classify.m`. The inputs of this function are training data, new data (either validation or testing data) and k . The function needs to output the accuracy on both training and new data (either validation or testing).
- Consider $k = 1, 3, 5, \dots, 23$. For each k , report the training and validation accuracy. Identify the k with the highest validation accuracy, and report the test accuracy with this choice of k . Note: if multiple values of k result in the highest validation accuracy, then report test accuracies for all such values of k .
- Apply k NN on the `boundary.mat` dataset which is a binary classification dataset with only two features. You need to run k NN with $k = 1, 5, 15, 20$ and examine the decision boundary. A simple way to visualize the decision boundary is to draw 10000 data points on a uniform 100×100 grid in the square $(x, y) \in [0, 1] \times [0, 1]$ and classify them using the k NN classifier. Then, plot the data points with different markers corresponding to different classes. Repeat this process for all k and discuss the smoothness of the decision boundaries as k increases.

5 Academic Integrity Form

The course Academic Integrity Form can be found here:

http://web.cs.ucla.edu/~ameet/teaching/fall15/cs260/cs260_academic_integrity_form.pdf.

Please read and sign it.

Submission

Please provide the following as part of your submission:

- Provide your answers to problems 1-3, 4c, 4d, and 5 in hardcopy. The papers need to be stapled and submitted at the beginning of class on the due date.
- Email all source code to Nikos (nikos.karianakis at gmail) by the due date. The only acceptable languages are MATLAB and Octave.