Homework #1

CS 260: Machine Learning Algorithms
Prof. Ameet Talwalkar
Due: 1/18/17, 10am

1 Warmups *[15 points]*

a. **Multivariable Calculus**: Consider $y = x\sin(z)e^{-x}$. What is the partial derivative of y with respect to x?

b. **Mean and Variance**: If the variance of a zero-mean random variable X is σ^2, what is the variance of $2X$? What about the variance of $X + 2$?

c. **Basic Probability**: Consider the following joint distribution between X and Y:

$P(X,Y)$		Y		
		a	b	c
X	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

What is $P(X = T|Y = b)$?

d. **Big-O notation**: For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, or both. Briefly justify your answers.

(a) $f(n) = ln(n), g(n) = lg(n)$ [ln denotes log to the base e and lg denotes log to the base 2.]

(b) $f(n) = 3^n, g(n) = n^{10}$

(c) $f(n) = 3^n, g(n) = 2^n$

e. **Divide and Conquer**: Assume that you are given an array with n elements all entries equal either to 0 or +1 such that all 0 entries appear before +1 entries. You need to find the index where the transition happens, *i.e.*, you need to report the index with the last occurrence of 0. Give an algorithm that runs in time $O(\log n)$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

2 Convex Functions and Information Theory *[20 points]*

a. Show that the function $f(x) = |x| + \exp(x)$ is convex.

b. Suppose the random variable X is distributed according to a k-class multi-nominal distributions with class probabilities $p_1, p_2, \ldots, p_k$, such that $\sum_{i=1}^{k} p_i = 1$. Find the values of $p_i, i = 1, \ldots, k$ such that the entropy of X is maximized.

3 Linear Algebra *[20 points]*

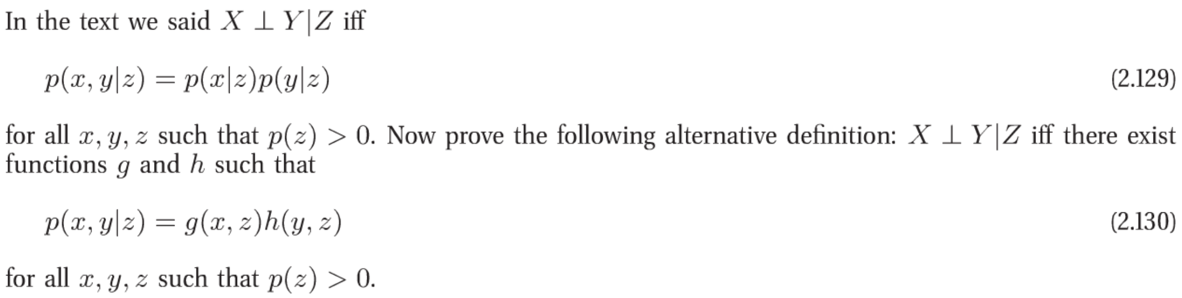
a. The covariance matrix $\boldsymbol{\Sigma}$ of a random vector $\boldsymbol{X}$ is defined as $\boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})^{\mathrm{T}}]$, where $\mathbb{E}\boldsymbol{X}$ is the expectation of $\boldsymbol{X}$. Is $\boldsymbol{\Sigma}$ positive-semidefinite?

b. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $\mathbb{R}^{\mathsf{D}\times\mathsf{D}}$ symmetric matrices. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ have the exact same set of eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{\mathsf{D}}$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \cdots, \alpha_{\mathsf{D}}$ for $\boldsymbol{A}$, and $\beta_1, \beta_2, \cdots, \beta_{\mathsf{D}}$ for $\boldsymbol{B}$. Please write down the eigenvectors and their corresponding eigenvalues for the following matrices:

- $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$
- $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$
- $\boldsymbol{E} = \boldsymbol{A}\boldsymbol{B}$
- $\boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}$ (assume $\boldsymbol{A}$ is invertible)

c. Let us assume $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive semidefinite. Show that the matrix $\boldsymbol{H} \in \mathbb{R}^{\mathsf{D}\times\mathsf{D}}$ is also positive semidefinite, where the ij-th element of $\boldsymbol{H}$ is defined as $H_{ij} = A_{ij}B_{ij}$. (Hint: Represent $\boldsymbol{H}$ using the eigendecompositions of $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e., $\boldsymbol{A} = \sum_d \alpha_d \boldsymbol{u}_d \boldsymbol{u}_d^{\top}$ and $\boldsymbol{B} = \sum_d \beta_d \boldsymbol{q}_d \boldsymbol{q}_d^{\top}$.)

4 Conditional Independence [from MLAPA] *[20 points]*



In the text we said $X \perp Y | Z$ iff

$$p(x, y|z) = p(x|z)p(y|z) \tag{2.129}$$

for all x, y, z such that $p(z) > 0$. Now prove the following alternative definition: $X \perp Y | Z$ iff there exist functions g and h such that

$$p(x, y|z) = g(x, z)h(y, z) \tag{2.130}$$

for all x, y, z such that $p(z) > 0$.

5 KNN Classification in MATLAB/Octave *[20 points]*

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.

a. 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 ith binary feature indicates whether the original feature belongs to the ith category or not. This strategy is called 'one-hot encoding.'

b. 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).

c. Consider $k = 1, 3, 5, \cdots, 23$. 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.

d. Apply kNN on the `boundary.mat` dataset which is a binary classification dataset with only two features. You need to run kNN 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 kNN 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.

6 Academic Integrity Form *[5 points]*

Please read and sign the course Academic Integrity Form:

http://web.cs.ucla.edu/~ameet/teaching/winter17/cs260/cs260_academic_integrity_form.pdf.

Submission Instructions

- Provide your answers to problems 1-4, 5c, 5d, and 6 in hardcopy. The papers need to be stapled, and submitted at the beginning of class on the due date.
- For problem 5, please put all of your code in a single folder named `[lastname]_[firstname]_hw1`, and submit a single `.zip` file containing this folder called `[lastname]_[firstname]_hw1.zip` to **CCLE** by the due date. Note: If you are not registered for the course yet, please email Brooke Wenig (brookewenig at gmail) to get access to CCLE.
- The only acceptable programming languages are MATLAB and Octave.
- This HW will NOT be graded unless you have taken the course Math Quiz.