Your Name
email@fas.harvard.edu
CS281-F17

Assignment #0, v 1.0
Due: 5:00pm Sept. 8th

Collaborators: John Doe, Fred Doe

Homework 0: Preliminary

Introduction

There is a mathematical component and a programming component to this homework. Please submit your PDF and Python files to Canvas, and push all of your work to your GitHub repository. If a question requires you to make any plots, please include those in the writeup.

This assignment is intended to ensure that you have the background required for CS281, and have studied the mathematical review notes provided in section. You should be able to answer the problems below *without* complicated calculations. All questions are worth $70/6 = 11.\bar{6}$ points unless stated otherwise.

Variance and Covariance

Problem 1

Let X and Y be two independent random variables.

(a) Show that the independence of X and Y implies that their covariance is zero.

(b) Zero covariance *does not* imply independence between two random variables. Give an example of this.

(c) For a scalar constant a, show the following two properties:

$$\mathbb{E}(X + aY) = \mathbb{E}(X) + a\mathbb{E}(Y)$$
$$\mathrm{var}(X + aY) = \mathrm{var}(X) + a^2\mathrm{var}(Y)$$

Densities

Problem 2

Answer the following questions:

(a) Can a probability density function (pdf) ever take values greater than 1?

(b) Let X be a univariate normally distributed random variable with mean 0 and variance $1/100$. What is the pdf of X?

(c) What is the value of this pdf at 0?

(d) What is the probability that $X = 0$?

(e) Explain the discrepancy.

Conditioning and Bayes' rule

Problem 3

Let $\mu \in \mathbb{R}^m$ and $\mathbf{\Sigma}, \mathbf{\Sigma}' \in \mathbb{R}^{m \times m}$. Let X be an m-dimensional random vector with $X \sim \mathcal{N}(\mu, \mathbf{\Sigma})$, and let Y be a m-dimensional random vector such that $Y \mid X \sim \mathcal{N}(X, \mathbf{\Sigma}')$. Derive the distribution and parameters for each of the following.

(a) The unconditional distribution of Y.

(b) The joint distribution for the pair (X, Y).

Hints:

- You may use without proof (but they are good advanced exercises) the closure properties of multivariate normal distributions. Why is it helpful to know when a distribution is normal?
- Review Eve's and Adam's Laws, linearity properties of expectation and variance, and Law of Total Covariance.

I can Ei-gen

Problem 4

Let $\mathbf{X} \in \mathbb{R}^{n \times m}$.

(a) What is the relationship between the n eigenvalues of $\mathbf{X}\mathbf{X}^T$ and the m eigenvalues of $\mathbf{X}^T\mathbf{X}$?

(b) Suppose $\mathbf{X}$ is square (i.e., $n = m$) and symmetric. What does this tell you about the eigenvalues of $\mathbf{X}$? What are the eigenvalues of $\mathbf{X} + \mathbf{I}$, where $\mathbf{I}$ is the identity matrix?

(c) Suppose $\mathbf{X}$ is square, symmetric, and invertible. What are the eigenvalues of $\mathbf{X}^{-1}$?

Hints:

- Make use of singular value decomposition and the properties of orthogonal matrices. Show your work.
- Review and make use of (but do not derive) the spectral theorem.

Vector Calculus

Problem 5

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{m \times m}$. Please derive from elementary scalar calculus the following useful properties. Write your final answers in vector notation.

(a) What is the gradient with respect to $\mathbf{x}$ of $\mathbf{x}^T\mathbf{y}$?

(b) What is the gradient with respect to $\mathbf{x}$ of $\mathbf{x}^T\mathbf{x}$?

(c) What is the gradient with respect to $\mathbf{x}$ of $\mathbf{x}^T\mathbf{A}\mathbf{x}$?

Gradient Check

Problem 6

Often after finishing an analytic derivation of a gradient, you will need to implement it in code. However, there may be mistakes - either in the derivation or in the implementation. This is particularly the case for gradients of multivariate functions.

One way to check your work is to numerically estimate the gradient and check it on a variety of inputs. For this problem we consider the simplest case of a univariate function and its derivative. For example, consider a function $f(x) : \mathbb{R} \to \mathbb{R}$:

$$\frac{df}{dx} = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x-\epsilon)}{2\epsilon}$$

A common check is to evaluate the right-hand side for a small value of ϵ, and check that the result is similar to your analytic result.

In this problem, you will implement the analytic and numerical derivatives of the function

$$f(x) = \cos(x) + x^2 + e^x.$$

1. Implement `f` in Python (feel free to use whatever `numpy` or `scipy` functions you need):

```python
def f(x):
```

2. Analytically derive the derivative of that function, and implement it in Python:

```python
def grad_f(x):
```

3. Now, implement a gradient check (the numerical approximation to the derivative), and by plotting, show that the numerical approximation approaches the analytic as `epsilon` $\to 0$ for a few values of x:

```python
def grad_check(x, epsilon):
```