CSCI-4150 Homework 0

This homework reviews math you should be comfortable with before taking the course. For each of the 16 problems, grade yourself as follows:

[Grade 0] *I have no idea what the problem is talking about.*

[Grade 1] *I understand what the problem is asking.*

[Grade 2] *I think I know how to solve after brushing up my knowledge.*

[Grade 3] *I can solve the problem.*

You do not have to solve any of the problems. Just honestly write a grade for each problem. Now, compute your average grade and write that on the front of your homework. Hand in this homework with the average as well as the individual grades for each problem.

Acknowledgements: the problems are designed by Malik Magdon-Ismail for CSCI-4100/6100 Machine Learning from Data.

Probability and Counting

1.(**Expectation**)
Random variables X, Y have expectations $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 2$. Let $Z = X + Y$.

(a) What is $\mathbb{E}[Z]$?

(b) How large can $\text{Var}[X]$ be, and how small can it be?

2.(**Binomial Distribution**)
You have 10 independent coins, each having probability $\frac{3}{5}$ of flipping heads. Let X be the number of heads flipped when all ten coins are flipped.

(a) What is $\mathbb{P}[X = 6]$?

(b) What is $\mathbb{P}[X \geq 6]$?

(c) What is $\mathbb{E}[X]$?

(d) What is $\text{Var}[X]$?

3.(**Mean and Variance**)
You have numbers $x_1, \ldots, x_N$. The mean is $\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$ and the variance is $\sigma_x^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})^2$.

(a) Show that $\sigma_x^2 = \frac{1}{N} \sum_{n=1}^{N} x_n^2 - \bar{x}^2$.

(b) Show that for any set of numbers, the average of the squares is as large as the square of the average.

(c) How can you compute the variance with just one pass through the data $x_1, \ldots, x_N$, using $O(1)$ memory.

4.(**Bounding Probabilities and Union Bound**)
Let A and B be arbitrary events. Suppose $\mathbb{P}[A] = 0.6$ and $\mathbb{P}[B] = 0.7$.

(a) What is the maximum possible value of $\mathbb{P}[A \cap B]$?

(b) What is the minimum possible value of $\mathbb{P}[A \cap B]$?

(c) What is the maximum possible value of $\mathbb{P}[A \cup B]$?

(d) What is the minimum possible value of $\mathbb{P}[A \cup B]$?

5.(**Random Variate Generation**)
You have a stream of independent uniform random variables taking on values in $[0, 1]$.

(a) How will you generate a uniform random variable in the interval $[a, b]$.

(b) How will you generate a Gaussian random variable with mean 0 and variance 1.

(c) How will you generate a Gaussian random vector $\mathbf{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

6.(**Combinatorics**)
You have a deck of 52 cards in which there are 4 cards of each denomination $1, 2, \ldots, 13$.

(a) In how many ways can you draw a full house, which is a hand of 5 cards with denominations $xxxyy$?

(b) What is the probability of obtaining a full house?

(c) In how many ways can you sample 10,000 balls from a set of 100,000 distinct balls. Give a decent estimate, not just ∞. (You may either numerically try to estimate this by (say) first estimating the logarithm of this number, or you may use an analytic approximation to the factorial function.)

7.(**Conditional Probability/Bayes Theorem**)
You roll two dice.

(a) What is the probability that at least one of the dice has rolled a 4?

(b) You are told the sum is X. Conditioned on the value of X, for $X = 2, 3, \ldots, 12$, what now is the probability that at least one of the dice has rolled a 4.

(c) Compute the probability that 'the sum is 10' conditioned on 'at least one of the dice has rolled a 4'.

Linear Algebra

8.(**Inner Product**)
The Euclidean inner (dot) product is $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{d} x_i y_i$.

(a) What is the formula for the norm, $\|\mathbf{x}\|$ and relate it to a dot product.

(b) What is the formula for the dot product in terms of the angle between the two vectors and the norms.

(c) Show that $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\mathbf{x} \cdot \mathbf{y}$.

(d) Show that $(\mathbf{x} \cdot \mathbf{y})^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$. When does equality occur?

(e) What is the maximum and minimum of $\mathbf{x} \cdot \mathbf{y}$ and $|\mathbf{x} \cdot \mathbf{y}|$, and give examples in 2 dimensions attaining these maxima and minima.

Multivariate Calculus

9.(**Derivatives**)

$$
\begin{aligned}
f(x) &= \ln(1 + e^{-2x}) \\
g(x, y) &= e^x + e^{y^2} + e^{3xy}
\end{aligned}
$$

What are $\dfrac{df}{dx}$ and $\dfrac{\partial g}{\partial y}$?

10.(**Chain Rule**)

$$
\begin{aligned}
f(u,v) &= uv \\
u(x,y) &= \cos(x+y) \\
v(x,y) &= \sin(x-y).
\end{aligned}
$$

What is $\dfrac{\partial f}{\partial x}$?

11.(**Integration**)

(a) Compute $\displaystyle\int_5^{10} dx \; \frac{x}{2x^2-3}$.

(b) Compute $\displaystyle\int_0^\infty dx \; \frac{1}{1+x^2}$.

12.(**Taylor Expansion**)

$$
E(u,v) = (ue^v - 2ve^{-u})^2.
$$

(a) Compute the gradient ∇E and the Hessian $\nabla^2 E$ evaluated at $u=1, v=1$.

(b) Give the second order Taylor series approximation to E at $u=1, v=1$.

13.(**Minimization**)

(a)
$$
E(w) = ae^w + be^{-2w},
$$
for constants $a, b > 0$. As a function of a, b, compute $\min_w E(w)$.

(b)
$$
E(x,y) = x^3 - 6xy + 3y^2 - 24x + 4
$$
Determine the stationary points of $E(x,y)$ and check for local minima or maxima.

14.(**Continuity**)
Consider
$$
f(x,y) = \frac{x^2 y}{x^4 + y^2} \qquad \text{for } (x,y) \neq (0,0) \text{ and zero otherwise.}
$$
Is f continuous at $(x,y) = (0,0)$? Why or why not?

15.(**Quadratic Forms**)
Given a matrix X and a vector $\mathbf{y}$ of appropriate dimensions, define
$$
E(\mathbf{w}) = \|X\mathbf{w} - \mathbf{y}\|^2.
$$

(a) If X is $N \times d$, what are the possible dimensions of $\mathbf{w}, \mathbf{y}$?

(b) Show that $E(\mathbf{w}) = \mathbf{w}^T X^T X \mathbf{w} - 2\mathbf{w}^T X^T \mathbf{y} + \mathbf{y}^T \mathbf{y}$.

(c) Show that the gradient and Hessian are given by:

$$
\begin{aligned}
\nabla E(\mathbf{w}) &= 2(X^T X)\mathbf{w} - 2X^T \mathbf{y} \\
\nabla^2 E(\mathbf{w}) &= X^T X.
\end{aligned}
$$

3

(d) If $X^T X$ is positive definite, show that $E(\mathbf{w})$ is minimized at $\mathbf{w}^* = (X^T X)^{-1} X^T \mathbf{y}$.

(e) What is $\min_{\mathbf{w}} E(\mathbf{w})$?

16.(**Limits**)

What is $\displaystyle\lim_{N \to \infty} \frac{e^N}{N + 2e^N}$?