Homework 0: review of probability, statistics, linear algebra, and calculus

CSCI145/MATH166, Mike Izbicki

DUE: Thursday, 12 September at the beginning of class

Name:

Problem 1. (10 pts) Section 1.2 of Bishop defines the following terms. Reproduce their definitions below.

1. sum rule

2. product rule

3. Bayes Theorem

4. expectation

5. linearity of expectation
   (not formally defined in Bishop, but is an immediate consequence of the definition of expectation)

$$\text{if } x, y \text{ are random variables, then } \mathbb{E}(x + y) = \mathbb{E}x + \mathbb{E}y$$

6. law of large numbers (Eq. 1.35)

7. conditional expectation

8. variance

9. covariance

10. standard deviation

11. mode

I will use the following terms frequently in class, so you need to know what they mean, but you do not need to be able to formally define them: *marginal distribution, conditional distribution, joint distribution, prior distribution, posterior distribution, normalization, probability density function (PDF), cumulative distribution function (CDF), probability mass function (PMF), independent and identically distributed, bias, overfitting, frequentist statistics versus bayesian statistics, likelihood function, error function, maximum likelihood (ML), maximum a posteriori (MAP), hyperparameters, measure theory.*

Problem 2. (15 pts) Prove or give a counter example: If two random variables x,y are independent, then their covariance is zero.

Problem 3. (15 pts) If two random variables x,y have zero covariance, then they are independent.

Problem 4. (10pts) The variance can be defined as either

$$\mathrm{Var}(x) = \mathbb{E}(x - \mathbb{E}(x))^2 \tag{1}$$

or

$$\mathrm{Var}(x) = \mathbb{E}(x^2) - (\mathbb{E}x)^2. \tag{2}$$

Show that these two definitions are equivalent. HINT: The problem requires applying the definition of expectation and algebraic manipulations.

Problem 5. (5 pts) Section 9.5 of *The Matrix Cookbook* defines orthogonal matrices and lists some of their properties. Reproduce this definition and properties below. (You do not need to reproduce subsections 9.5.1, 9.5.2, or 9.5.3, just the main information under 9.5.)

Problem 6. (5 pts) Sections 9.6.1-9.6.5 of *The Matrix Cookbook* defines positive (semi-)definite matrices and lists some of their properties. Reproduce this definition and properties below.

Problem 7. (10 pts) A vector/matrix *norm* must satisfy four requirements (Equations 529-532 of *The Matrix Cookbook*). State them.

1.

2.

3.

4.

Define the L_1 vector norm (Eq. 525), L_2 vector norm (Eq. 526), induced/operator norm of a matrix, the L_1 matrix norm (Eq. 537), the L_2 matrix norm (Eq. 538), and the Frobenius norm (Eq. 541).

Problem 8. (15 pts) The log-loss is defined to be

$$\ell(\mathbf{x};\mathbf{w}) = \log(1 + \exp(-\mathbf{x}^T\mathbf{w})). \tag{3}$$

Calculate the first and second derivatives of (3) with respect to $\mathbf{w}$. Note that because $\mathbf{w}$ is a vector, the loss is a scalar, the derivative is a vector, and the second derivative is a matrix. See Section 2 of *The Matrix Cookbook* for a review of vector derivatives.

Problem 9. (15 pts) Calculate

$$\arg\min_{\mathbf{w}} \left(\|\mathbf{y} - X\mathbf{w}\|^2 + \lambda\|\mathbf{w}\|^2 \right) \tag{4}$$

where $\mathbf{w}$ is a vector of dimension d, $\mathbf{y}$ is a vector of dimension n, and X is a matrix with shape $n \times d$.

HINT: Recall that the arg min is defined as

$$\arg\min_{\mathbf{w}} f(\mathbf{w}) \triangleq \{\mathbf{w} : f(\mathbf{w}) = \min_{\mathbf{w}'} f(\mathbf{w}')\}. \tag{5}$$

That is, the arg min returns the set of values that minimize the function f. To calculate the arg min, take the derivative of f with respect to $\mathbf{w}$, set it equal to zero, and solve for $\mathbf{w}$.

Problem 10. (2pts extra credit) If you complete this assignment in LaTeX, then you will receive +2 pts extra credit. The source files are available at:

`https://github.com/mikeizbicki/cmc-csci145-math166/tree/master/hw_00`.