# 6.867 Fall 2017 : Homework 0

This is a set of warm-up problems, divided into two parts. *There is no need to turn in this homework. We will post solutions.*

Problems in the Part 1 constitute a quick diagnostic, to see if your background in linear algebra and probability is adequate to do well in the course. If these problems are not fairly easy for you, then you should probably take the relevant courses this term and take 6.867 next year.

Problems in the second part are intended to help you learn good "vectorized" programming strategies, which will help you make efficient implementations of algorithms we study in class and also to interface with existing libraries in MATLAB and Python. We don't expect you to know how to do this already, but ask you to spend some time in the first week learning how to construct elegant, efficient solutions to these problems. You are welcome to use MATLAB or numpy/scipy in Python.

## 1 Math background problems

### 1.1 Just plane fun

Consider a hyperplane in $n$-dimensional Euclidean space, described by the $n+1$ real values $w_i$ for $i = 0, \ldots, n$: the hyperplane consists of points $(x_1, \ldots, x_n)$ satisfying

$$w_0 + w_1 x_1 + \ldots + w_n x_n = 0 \ .$$

1. Find a unit vector normal to the hyperplane. Given a point $\boldsymbol{v} = (v_1, \ldots, v_n)$ on the hyperplane, give the equation for the line through the point that is orthogonal to the hyperplane.

2. Given a point $\boldsymbol{v} = (v_1, \ldots, v_n)$, how can you determine which side of the hyperplane it is on?

3. What is the distance of a point $\boldsymbol{v} = (v_1, \ldots, v_n)$ to the hyperplane?

4. Consider the halfspace $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{w}^T \boldsymbol{x} \leq b\}$. Find a nearest (in Euclidean distance) point in $H$ to a given vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$. Is this point unique? Hint: Consider combining the Euclidean distance and the constraint into a single function to optimize by using a Lagrangian multiplier.

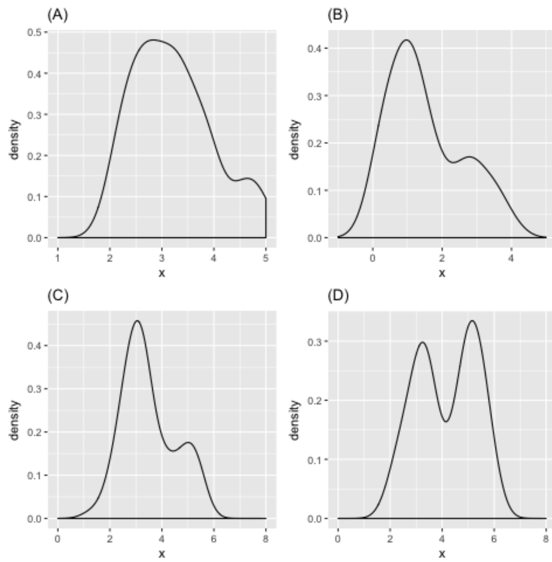
### 1.2 Multivariate Gaussian

Let $X$ be a random variable taking values in $\mathbb{R}^n$. It is normally distributed with mean $\mu$ and covariance matrix $\Sigma$.

1. Write the probability density function (pdf) $p_X(\boldsymbol{x})$, sometimes denoted $p(X = \boldsymbol{x})$, for $X$.
2. Show how to obtain the normalization constant for the multivariate Gaussian.
3. Let $Y = 2X$. What is the pdf of $Y$?
4. What can we say about the distribution of $X$ if $\Sigma$ is the identity matrix, $I$? Does this imply anything about factorization of the pdf?
5. What can we say about the distribution of $X$ if $\Sigma$ is $5I$?
6. What can we say about the distribution of $X$ if $\Sigma$ is $((10, 0), (0, 1))$?
7. Approximately what shape do equiprobability contours (i.e., sets $\{x \in R^n : p_X(x) = c\}$ for some $c$) of this distribution have?
8. What can we say about this distribution if $\Sigma$ is $((10, -4), (-4, 10))$?
9. Approximately what shape do equiprobability contours of this distribution have?
10. Is $((2, 10), (10, 2))$ a valid $\Sigma$? How can you tell?
11. If $n = 2$, $\mu = (1, 2)$, and $\Sigma = ((10, 0), (0, 1))$, what is the conditional pdf $p_{X_1|X_2}(x_1 \mid 3)$?

### 1.3 Probability

1. You go for your annual checkup and have several lab tests performed. A week later your doctor calls you and says she has good and bad news. The bad news is that you tested positive for a marker of a serious disease, and that the test is 97% accurate (i.e. the probability of testing positive given that you have the disease is 0.97, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only 1 in 20,000 people. What are the chances that you actually have the disease?
2. Consider the following generative process describing the joint distribution $p(Z, X) : Z \sim$ Bernoulli$(0.2), X \sim$ Gaussian$(\mu_Z, 0.5)$, where $\mu_0 = 3$ and $\mu_1 = 5$. Which of the following plots is the marginal distribution $p(X)$?




### 1.4 Multivariate calculus

1. Find the minimum value of the function $f(x, y) = x^2 + 2y^2 - xy + x - 4y$ over $\mathbb{R}^2$.
2. Prove that $f(x, y)$ is convex over $\mathbb{R}^2$ by showing that its Hessian is positive semi-definite.

## 2 Programming problems

Try your best to avoid writing your own `for` loops!! (i.e. “vectorized” programming)

### 2.1 Regularization

Given an $n \times n$ matrix $C$, add a scalar $a$ to each diagonal entry of $C$.

### 2.2 Largest Off-diagonal Element

Given an $n \times n$ matrix $A$, find the largest off-diagonal element.

### 2.3 Pairwise Computation

Given a vector $x$ of length $m$, and a vector $y$ of length $n$, compute $m \times n$ matrices: $A$ and $B$, such that $A(i, j) = x(i) + y(j)$, and $B(i, j) = x(i) \cdot y(j)$.

### 2.4 Pairwise Euclidean Distances

Given a $d \times m$ matrix $X$, and a $d \times n$ matrix $Y$, compute an $m \times n$ matrix $D$, such that $D(i, j) = \|x^i - y^j\|^2$, where $x^i$ is the $i$-th column of $X$, and $y^j$ is the $j$-th column of $Y$.

### 2.5 Compute Mahalanobis Distances

The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. This distance is zero if P is at the mean of D, and grows as P moves away from the mean: Along each principal component axis, it measures the number of standard deviations from P to the mean of D. If each of these axes is rescaled to have unit variance, then Mahalanobis distance corresponds to standard Euclidean distance in the transformed space. Mahalanobis distance is thus unitless and scale-invariant, and takes into account the correlations of the data set (from `http://en.wikipedia.org/wiki/Mahalanobis_distance`). Given a center vector $c$, a covariance matrix $S$, and a set of $n$ vectors as columns in matrix $X$, compute the distances of each column in $X$ to $c$, using the following formula:

$$D(i) = (x^i - c)^T S^{-1} (x^i - c). \tag{4}$$

Here, $D$ is a row vector of length $n$.

### 2.6 2-D Gaussian

Generate 1000 random points from a 2-D Gaussian distribution with mean $\mu = [4, 2]$ and covariance

$$\Sigma = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 3 \end{pmatrix} \tag{5}$$

Plot the points so obtained, and estimate their mean and covariance from the data. Find the eigenvectors of the covariance matrix and plot them centered at the sample mean.