CSE – 426 Homework 0

Griffin Kent

The goal of this homework is to decide if you have enough background in probability, statistics, linear algebra, calculus, algorithms and programming, to take this math-intensive class. If you can accomplish both Sections (Minimum/Medium Background Tests) without too much effort, then you have a solid required background. If you can finish the questions in Section 1 but find those in Section 2 difficult, you can still take this course, but you will need to spend some more time outside the class to take care of those concepts and techniques that you're less familiar with. If you find Section 1 difficult, then you need to learn the prerequisites and come back next year.

Grading: This one will not be graded. Be honest to yourself: if you get solution from others without knowing the background, you are hiding away the issues that will come out again and again in the class.

Submitting: Only electronic submissions on Coursesite are accepted. You can handwrite your answers and then scan and upload them to Coursesite. The plotted files should be saved as pdf files. Then zip your scanned answers and the plots in pdf files into a single zip file, named as

<Your LIN>HW0.zip

It is your responsibility to name the submission correctly, since we will download the files in batch and your LIN will link the submissions to your account on our record.

Topics covered in this homework include:

1. **Linear algebra:** vectors, matrices, norms, rank, determinant, inverse and transposition.

2. **Calculus:** (partial) derivative of a scalar function with respect to another scalar or a vector; logarithm; exponential; integration.

3. **Probability:** continuous and discrete random variables and vectors, events, mean, variance, independence, conditional probability.

4. **Statistics:** find likelihood, mean and variance from observed data.

5. **Algorithm and programming:** understand the space and time complexity of algorithms; Python.

Here are some useful resources for you to refer to:

- **Probability**
  Lecture notes:
  http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf
  Chapter 3 of the Deep Learning book:
  https://www.deeplearningbook.org/contents/prob.html

• **Linear Algebra**
Short video lectures by Prof. Zico Kolter:
http://www.cs.cmu.edu/~zkolter/course/linalg/outline.html
Handout associated with above video:
www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf
Short video lectures by Prof. Andrew Ng:
https://www.coursera.org/lecture/machine-learning/matrices-and-vectors-38jIT
Chapter 2 of the Deep Learning book:
https://www.deeplearningbook.org/contents/linear_algebra.html
Matrix Cookbook:
https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

• **Python, numpy, scipy, plotting, IPython Notebook, etc.**
Stanford tutorial:
http://cs231n.github.io/python-numpy-tutorial/

1 Minimum Background Test

Linear Algebra

Consider the matrix X and the vectors y and z below:

$$X = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}, y = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, z = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

1) What is the inner product of the vectors y and z? (This is also sometimes called the dot product and is sometimes written $y^T z$)

(Solution): $y^T z = \begin{bmatrix} 3 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} = 3(6) + 5(1) = 18 + 5 = 23$. ■

2) What is the product Xy?

(Solution): $Xy = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 4(3) + 3(5) \\ 1(3) + 2(5) \end{bmatrix} = \begin{bmatrix} 27 \\ 13 \end{bmatrix}$. ■

3) Is X invertible? If, give the inverse, and if not, explain why.

(Solution): A matrix is invertible *iff* the determinant of the matrix is nonzero. Thus, calculating the determinant:

$$\det(x) = 4(2) - 3(1) = 5.$$

Therefore, X is invertible and we can find the inverse to be

$$X^{-1} = \frac{1}{\det(x)} \begin{bmatrix} 2 & -3 \\ -1 & 4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 2 & -3 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} \end{bmatrix}. ■$$

4) What is the rank of X?

(Solution) The rank of a matrix is the number of linearly independent column vectors or row vectors that make up the matrix. Since neither of the column vectors that make up X, $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$, can form the other by using a linear transformation, matrix X has rank 2.■

Calculus

1) If $y = \frac{1}{2}x^3 - 3x + 4$, then what is the derivative of y with respect to x?

(Solution):

$$\frac{dy}{dx} = \frac{3}{2}x^2 - 3. \blacksquare$$

2) If $y = 2x \cos(z) e^{-x}$, then what is the partial derivative of y with respect to x?

(Solution): Letting $f(x) = 2x \cos(z)$ and $g(x) = e^{-x}$, we can use the product rule

$$\frac{\partial y}{\partial x} = 2\cos(z)\, e^{-x} + 2x\, cos(z) e^{-x}(-1)$$
$$= 2\cos(z)\, e^{-x} - 2x\, cos(z) e^{-x}. \blacksquare$$

Probability and Statistics

Consider a sample of data $S = \{1,0,0,1,0\}$ created by flipping a coin five times and let the random variable $X = 0$ when the coin turned up heads and $X = 1$ denote that it turned up tails.

1) What is the sample mean of this data?

(Solution): The sample mean is calculated at follows

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1+0+0+1+0}{5} = \frac{2}{5}. \blacksquare$$

2) What is the sample variance for this data?

(Solution): The sample variance is calculated as follows

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
$$= \frac{\left(1-\frac{2}{5}\right)^2 + \left(0-\frac{2}{5}\right)^2 + \left(0-\frac{2}{5}\right)^2 + \left(1-\frac{2}{5}\right)^2 + \left(0-\frac{2}{5}\right)^2}{5-1}$$
$$\frac{2\left(\frac{3}{5}\right)^2 + 3\left(-\frac{2}{5}\right)^2}{4} = \frac{\frac{18}{25} + \frac{12}{25}}{4} = \frac{\frac{30}{25}}{4} = \frac{3}{10}.$$

Note*: Do not confuse the sample variance with the population variance which is defined as follows

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 . \blacksquare$$

3) What is the probability of observing this data, assuming it was generated by flipping a coin with an equal probability of heads and tails (i.e. the probability distribution is $p(X = 1) = 0.5$ and $p(X = 0) = 0.5$).

(Solution): The probability of generating a series of data with independent probability frequencies can be obtained to be

$$p(S) = (0.5)^5 = 0.03125. \blacksquare$$

4) Note that the probability of this data sample would be greater if the value of $p(X = 1)$ was not 0.5, but instead some other value. What is the value that maximizes the probability of the sample S. Please justify your answer.

(Solution): The probability of obtaining the sample S is maximized when the probabilities of $p(X = 0)$ and $p(X = 1)$ mirror the ratio that is found in the sample. Thus, since there are two instances of $X = 1$ and three instances of $X = 0$, the respective ratios are $\frac{2}{5}$ and $\frac{3}{5}$. Therefore, the probability of obtaining S will be maximized when $p(X = 1) = \frac{2}{5}$ and $p(X = 0) = \frac{3}{5}$, yielding $p(S) = \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^3 = 0.03456. \blacksquare$

5) Consider the following joint probability table over variables y and z, where y takes a value from the set $\{a, b, c\}$, and z takes a value from the set $\{T, F\}$.

	A	B	C
T	0.2	0.1	0.2
F	0.05	0.15	0.3

• What is $p(z = F\ AND\ y = c)$?

(Solution): We can use the table to determine where these two values intersect and see that

$$p(z = F, y = c) = 0.3. \blacksquare$$

• What is $p(z = F|y = a)$?

(Solution): Using the product rule, we can see that

$$p(z = F, y = a) = p(z = F|y = a)p(y = a)$$

$$\longrightarrow p(z = F|y = a) = \frac{p(z = F, y = a)}{p(y = a)}$$

$$= \frac{0.05}{0.25} = 0.2. \blacksquare$$

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.

1) $f(n) = \ln(n)$, $g(n) = \log(n)$. Note that ln denotes logarithm to the base e and log denotes logarithm to the base 2.

(Solution): Using $\lim_{n\to\infty}\frac{f(n)}{g(n)} = c$, we have $\lim_{n\to\infty}\frac{\ln(n)}{\log(n)}$. Using L'Hopital's rule to differentiate both the numerator and denominator, we have

$$\lim_{n\to\infty}\frac{\frac{1}{n}}{\frac{1}{n\ln(2)}} = \ln(2).$$

Since $0 < c = \ln(2) < \infty$, we know that $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Therefore, $f(n) = \Theta(g(n))$, or in other words, Both.■

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

(Solution): Using $\lim_{n\to\infty}\frac{f(n)}{g(n)} = c$, we have $\lim_{n\to\infty}\frac{3^n}{n^3}$. Using L'Hopital's rule to differentiate both the numerator and denominator, we have

$$\lim_{n\to\infty}\frac{3^n\ln(3)}{3n^2} = \lim_{n\to\infty}\frac{3^n(\ln(3))^2}{6n} = \lim_{n\to\infty}\frac{3^n(\ln(3))^3}{6} = \infty.$$

Since $0 < c = \infty$, we know that $f(n)$ is increasing faster than $g(x)$ and thus $f(n) = \Omega(g(n))$, or in other words $g(n) = O(f(n))$. ■

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

(Solution): Using $\lim_{n\to\infty}\frac{f(n)}{g(n)} = c$, we have $\lim_{n\to\infty}\frac{3^n}{2^n}$. Using L'Hopital's rule to differentiate both the numerator and denominator, we have

$$\lim_{n\to\infty}\frac{3^n}{2^n} = \infty.$$

Since $0 < c = \infty$, we know that $f(n)$ is increasing faster than $g(x)$ and thus $f(n) = \Omega(g(n))$, or in other words $g(n) = O(f(n))$. ■

2 Medium Background Test

Probability and Statistics

1) True or False. Suppose X and Y are two continuous random variables. We use $\mathbb{E}$ and Var to denote expectation and variance of a random variable and a is a real number.

(A) $\mathbb{E}[aX] = a\mathbb{E}[X]$. True.

(B) $Var[aX] = aVar[X]$. False; $Var[aX] = a^2Var[X]$.

(C) $Var[X+Y] = Var[X] + Var[Y]$. False; $Var[X+Y] = Var[X] + Var[Y] + 2cov(X,Y)$.

(D) $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. True.

(E) If X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. True.

2) Define the covariance of two random variables X and Y to be

$$cov(X,Y) = \mathbb{E}_{XY}[(X - \mathbb{E}_X[X])(Y - \mathbb{E}_Y[Y])].$$

Prove that $cov(X,Y) = \mathbb{E}_{XY}[XY] - \mathbb{E}_X[X]\mathbb{E}_Y[Y]$.

(Proof): We have

$$\begin{aligned} cov(X,Y) &= \mathbb{E}_{XY}[(X - \mathbb{E}_X[X])(Y - \mathbb{E}_Y[Y])] \\ &= \mathbb{E}_{XY}\big[XY - X\mathbb{E}_Y[Y] - Y\mathbb{E}_X[X] + \mathbb{E}_X[X]\mathbb{E}_Y[Y]\big] \\ &= \mathbb{E}_{XY}[XY] - \mathbb{E}_X[X]\mathbb{E}_Y[Y] - \mathbb{E}_Y[Y]\mathbb{E}_X[X] + \mathbb{E}_X[X]\mathbb{E}_Y[Y] \\ &= \mathbb{E}_{XY}[XY] - 2\mathbb{E}_X[X]\mathbb{E}_Y[Y] + \mathbb{E}_X[X]\mathbb{E}_Y[Y] \\ &= \mathbb{E}_{XY}[XY] - \mathbb{E}_X[X]\mathbb{E}_Y[Y]. \end{aligned}$$

■

3) Write down the probability density function (PDF) of an n-dimensional Gaussian random vector $\boldsymbol{x} \in \mathbb{R}^n$ with mean μ and covariance matrix Σ. ($\mathbb{R}^n$ means the set of all real vectors of length n).

(Solution): The Gaussian distribution defined over an n-dimensional vector $x \in \mathbb{R}^n$ is defined as the following

$$\mathcal{N}(\boldsymbol{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}}\frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\right\}. \blacksquare$$

Linear Algebra

1) Show that the inverse of an invertible and symmetric matrix is also symmetric. That is, if A is symmetric and invertible with the inverse A^{-1}, show that $(A^{-1})^T = A^{-1}$.

(Proof): Given a matrix A that is symmetric and invertible, then we know that

$$A^T = A$$

Taking the inverse of this, we obtain

$$(A^T)^{-1} = A^{-1}$$

Using the property of invertible matrices, we obtain the result

$$(A^{-1})^T = A^{-1}$$

Therefore, the inverse of an invertible and symmetric matrix is also symmetric.

■

2) Given the fact that the determinant of any matrix A (denoted by $|A|$) is the product of A's eigenvalues and $|AB| = |A||B|$. Show that A^{-1} has determinant $\frac{1}{|A|}$.

(Proof): By the definition of an invertible matrix A, we have

$$AA^{-1} = I$$

Taking the determinant of both sides and using the knowledge given above, we have

$$\begin{aligned} &|AA^{-1}| = |I| \\ &\longrightarrow |A||A^{-1}| = 1 \\ &\longrightarrow |A^{-1}| = \frac{1}{|A|} \end{aligned}$$

Therefore, completing the proof.

■

3) Given a vector $x \in \mathbb{R}^n$, prove that $\|x\|_2 \leq \|x\|_1$. ($\|x\|_p$ is the p-norm of x).

(Proof): First, let's look at the definitions of the 1-norm and 2-norm

$$\|x\|_1 \coloneqq \sum_{i=1}^{n} |x_i| \qquad \|x\|_2 \coloneqq \sqrt{\sum_{i=1}^{n} x_i^2}$$

We can restate what we are trying to prove as the following:

$$\sqrt{\sum_{i=1}^{n} x_i^2} \leq \sum_{i=1}^{n} |x_i| \longrightarrow \sum_{i=1}^{n} x_i^2 \leq \left(\sum_{i=1}^{n} |x_i|\right)^2$$

We will now use a proof by induction.

Base Case: For $n = 1$, x is just a scalar and it is trivial to see that

$$\|x\|_1 \geq \|x\|_2$$

Hypothesis: Let's assume that $\sum_{i=1}^{n} x_i^2 \leq (\sum_{i=1}^{n} |x_i|)^2$.

Induction: Now, we must show that $\sum_{i=1}^{n+1} x_i^2 \leq (\sum_{i=1}^{n+1} |x_i|)^2, \forall n > 1$.

To that end,

$$\left(\sum_{i=1}^{n+1} |x_i|\right)^2 = \left(\sum_{i=1}^{n} |x_i| + |x_{n+1}|\right)^2$$

$$\left(\sum_{i=1}^{n} |x_i|\right)^2 + 2\left(\sum_{i=1}^{n} |x_i|\right)|x_{n+1}| + |x_{n+1}|^2$$

$$= \left(\sum_{i=1}^{n} |x_i|\right)^2 + x_{n+1}^2 + 2\left(\sum_{i=1}^{n} |x_i|\right)|x_{n+1}| \geq \sum_{i=1}^{n} x_i^2 + x_{n+1}^2.$$

Therefore, we have proven that $\sum_{i=1}^{n} x_i^2 \leq (\sum_{i=1}^{n} |x_i|)^2, \forall n > 0$.

Taking the square-root of both sides, we can see that $\|x\|_2 \leq \|x\|_1$.

■

4) Given orthonormal basis vectors $\{u_1, \dots, u_n\}$ for the n-dimensional vector space, any vector $\boldsymbol{x} \in \mathbb{R}^n$ can be represented by $\sum_{i=1}^{n} x_i u_i$. Prove that $x_i = \boldsymbol{x}^T u_i$.

(Proof): By definition, a **Basis** of a set of vectors is simply the minimum set of linearly independent column vectors of that set. Before we begin, let's define the set of orthonormal basis vectors as the following matrix

$$U = [u_1, \dots, u_n]$$

Where u_i defines the ith column of matrix U. Now, to begin, we can see that

$$\boldsymbol{x}^T U = [x_1, x_2, \dots, x_n] \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix} = \begin{bmatrix} x_1 u_{11} + x_2 u_{21} + \cdots + x_n u_{n1} \\ x_1 u_{12} + x_2 u_{22} + \cdots + x_n u_{n2} \\ \vdots \\ x_1 u_{1n} + x_2 u_{2n} + \cdots + x_n u_{nn} \end{bmatrix}^T$$

As we can see, each row of this column vector is equal to $\sum_{i=1}^{n} x_i u_i$, which we know by the definition given above for orthonormal basis vectors, represents each element x_i of the vector $\boldsymbol{x}$. Therefore, we can see that $x_i = \boldsymbol{x}^T u_i$, $\forall i \in n$. ■

Calculus

1) Take the natural logarithm (ln) of the Gaussian PDF you obtained from question 3 in "Probability and Statistics" in section 2.

(Solution): The Gaussian PDF is defined as

$$\mathcal{N}(\boldsymbol{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}}\frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\right\}$$

Taking the natural logarithm, we have

$$\begin{aligned}\ln\big(\mathcal{N}(\boldsymbol{x}|\mu,\Sigma)\big) &= \ln\left(\frac{1}{(2\pi)^{n/2}}\right) + \ln\left(\frac{1}{|\Sigma|^{\frac{1}{2}}}\right) - \frac{1}{2}\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big)\\ &= 0 - (2\pi)^{\frac{n}{2}} + 0 - |\Sigma|^{\frac{1}{2}} - \frac{1}{2}\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big)\\ &= -(2\pi)^{\frac{n}{2}} - |\Sigma|^{\frac{1}{2}} - \frac{1}{2}\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big). \blacksquare\end{aligned}$$

2) Take the derivative of the natural logarithm of the Gaussian PDF with respect to the mean μ. (Hint: you will need to learn about the trace of a matrix, and then use Eq. (108) in the Matrix Cookbook).

(Solution): Working from what we have above,

$$\begin{aligned}\frac{d}{d\mu}\Big(\ln\big(\mathcal{N}(\boldsymbol{x}|\mu,\Sigma)\big)\Big) &= \frac{d}{d\mu}\left(-(2\pi)^{\frac{n}{2}} - |\Sigma|^{\frac{1}{2}} - \frac{1}{2}\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big)\right)\\ &= -\frac{1}{2}\frac{d}{d\mu}\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big)\\ &= -\frac{1}{2}\frac{d}{d\mu}Tr\big((\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\big)\\ &= -\frac{1}{2}\big(\Sigma^{-1}(\boldsymbol{x}-\mu) + \Sigma^{-T}(\boldsymbol{x}-\mu)\big)\\ &= \frac{1}{2}\big(\Sigma^{-1}(\boldsymbol{x}-\mu) - \Sigma^{-T}(\boldsymbol{x}-\mu)\big). \blacksquare\end{aligned}$$

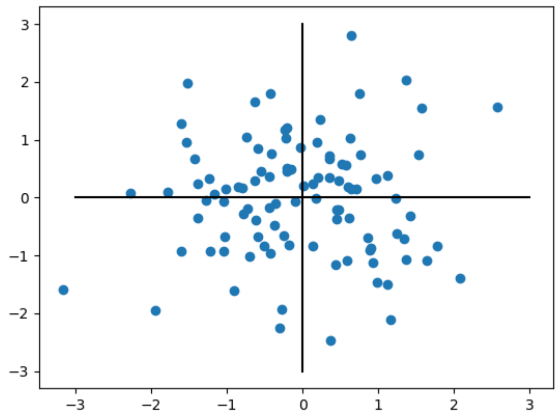
Numpy and Plotting

Depending on your OS, follow the links in the following link to set up Anaconda on your computer:
https://docs.anaconda.com/anaconda/install/
Then use Python, Numpy, and Matplotlib associated with Anaconda to plot some figures and to understand multi-dimensional Gaussian distributions.

1) Using Numpy to sample 100 points from the Gaussian distribution with mean $(0,0)$ and covariance $\Sigma = I_{2\times 2}$ (2x2 identity matrix). Plot the points.

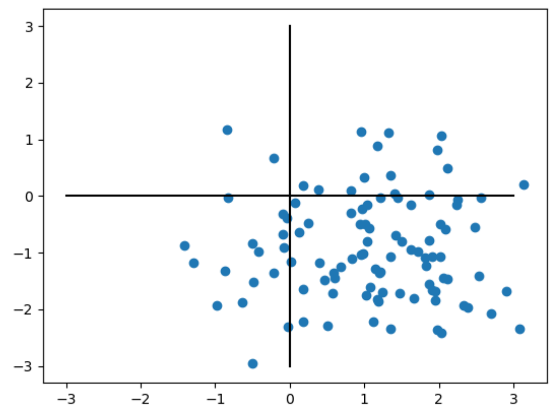
(Solution): We can use the following code snip to generate a sample from the 2-dimensional Gaussian distribution and plot it.

```python
import numpy as np
import matplotlib.pyplot as plt
n_samples = 100
mean = [0, 0]
covariance = [[1, 0], [0, 1]]
x, y = np.random.multivariate_normal(mean, covariance, n_samples).T
plt.plot(x, y, 'o')
plt.plot([-3,3], [0,0], color='black')
plt.plot([0,0], [-3,3], color='black')
```



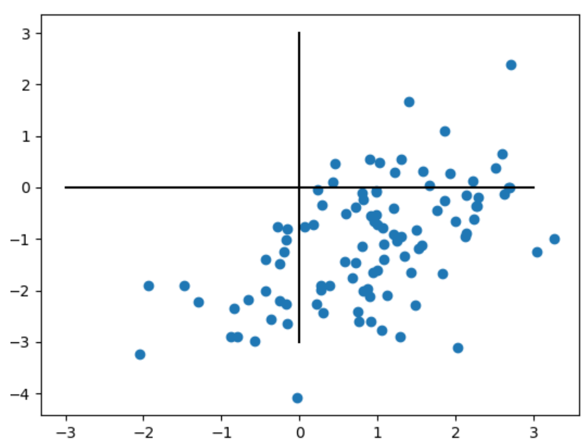
2) Now change the mean to $(1, -1)$ and plot the sample again.

(Solution):



3) Now sample and plot using $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

(Solution):



4) Now sample and plot using $\Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$.

(Solution):