CS188, Winter 2017
Problem Set 0: Math Prerequisites
Due Jan 19,2017

1 Problem 1

Solution: Solution to problem 1a

$$y' = sin(z)(e^{-x})(1-x)$$

2 Problem 2

(a) Problem 2a **Solution:** Solution to problem 2a

$$\begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 + 9 = 11$$

(b) Problem 2b **Solution:** Solution to problem 2b

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \text{ x } 1 + 3 \text{ x } 4 \\ 1 \text{ x } 1 + 3 \text{ x } 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$$

(c) Problem 2c **Solution:** Solution to problem 2c

$$\frac{1}{2*3-4*1} \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1.5 & -2 \\ -0.5 & 1 \end{bmatrix}$$

(d) Problem 2d **Solution:** Solution to problem 2d

Since the matrix is invertible, X is full rank and therefore has rank 2.

3 Problem 3

(a) Problem 3a **Solution:** Solution to problem 3a

$$\frac{1+1+0+1+0}{5} = \frac{3}{5}$$

(b) Problem 3b **Solution:** Solution to problem 3b

$$Var[X] = \frac{3}{5} - (\frac{3}{5})^2 = \frac{6}{25}$$

(c) Problem 3c **Solution:** Solution to problem 3c

$$\frac{1}{2^5} = \frac{1}{32}$$

(d) Problem 3d **Solution:** Solution to problem 3d

Let p be the probability of getting heads and $f(p)$ be the probability of the occurrence of the sample set S. We use critical points to find the value p that maximizes f(p).

$$f(p) = (p)(p)(1-p)(p)(1-p) = p^5 - 2p^4 + p^3$$

$$f'(p) = 5p^4 - 8p^3 + 3p^2$$

Critical points (given by factoring and quadratic formula): $0, 0, 3/5, 1$

$$f(0) = 0, f(3/5) = \frac{108}{3125} = 0.03456, f(1) = 0$$

The value that maximizes the probability is $\frac{3}{5}$.

(e) Problem 3e **Solution:** Solution to problem 3e

$$\frac{0.1}{0.1 + 0.15} = \frac{0.1}{0.25} = 0.4$$

4 Problem 4

(a) Problem 4a **Solution:** Solution to problem 4a

False

(b) Problem 4b **Solution:** Solution to problem 4b

True

(c) Problem 4c **Solution:** Solution to problem 4c

False

(d) Problem 4d **Solution:** Solution to problem 4d

False

(e) Problem 4e **Solution:** Solution to problem 4e

True

5 Problem 5

(a) Problem 5a **Solution:** Solution to problem 5a

v

(b) Problem 5b **Solution:** Solution to problem 5b

iv

(c) Problem 5c **Solution:** Solution to problem 5c

ii

(d) Problem 5d **Solution:** Solution to problem 5d

i

(e) Problem 5e **Solution:** Solution to problem 5e

iii

6 Problem 6

(a) Problem 6a **Solution:** Solution to problem 6a

$$Mean : E[X] = (0)(0.5) + (1)(0.5) = 0.5$$
$$Variance : Var[X] = 0.5 - 0.5^2 = 0.25$$

(b) Problem 6b **Solution:** Solution to problem 6b

$$Var[2X] = 4\sigma^2$$

$$Var[X+2] = E[X^2 + 4X + 4] - E[X+2]^2$$
$$Var[X+2] = E[X^2] + E[4X] + E[4] - (E[X]^2 + 2E[X]E[2] + E[2]^2)$$
$$Var[X+2] = E[X^2] - E[X]^2 + 4E[X] - 4E[X] + 4 - 4$$
$$Var[X+2] = E[X^2] - E[X]^2 = \sigma^2$$

7 Problem 7

(a) Problem 7a **Solution:** Solution to problem 7a

i. We know that $ln(2)lg(n) = ln(n)$, and $ln(2)$ is a constant factor between $ln(n)$ and $lg(n)$, which means that both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ are valid.

ii. $g(n) = O(f(n))$, since $f(n)$ is an exponential function and grows faster than $g(n)$, which is a polynomial function.

iii. $g(n) = O(f(n))$, since the $f(n)$ is grows more quickly than $g(n)$ by a factor of$(\frac{3}{2})^n$, which is not a constant factor, meaning that $f(n)$ will always be an upper bound of $g(n)$ (for sufficiently large n) and not vice-versa.

(b) Problem 7b **Solution:** Solution to problem 7b

Algorithm description: Building upon a binary search algorithm, start at the middle of the list jump left-wards when encountering a 1 and rightwards when encountering at 0. Return the final index if the item is 0 or (index - 1) if the item is +1.

Proof of Correctness: In order to find the transition point, we want to eliminate all elements of the array that could not be the transition point: namely, all indices on the left of 0's and all indices on the right of 1's. This is achieved in the binary search fashion. Eventually, the search space is reduced to a list of a single item (whether this is a 0 or +1 will be implementation based), which will indicate the transition point.

Runtime Complexity Justification: Binary search is $O(logN)$, and since this algorithm does trivial (constant time) operations on top of the classic binary search, the runtime complexity is the same.

8 Problem 8

(a) Problem 8a **Solution:** Solution to problem 8a

i.

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dydx$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dydx$$

$$E[XY] = \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy$$
$$E[XY] = E[X]E[Y]$$

(b) Problem 8b **Solution:** Solution to problem 8b

i. By the law of large numbers, the frequency of a particular event occurring will converge with the expected value given by the probability of it occurring, which in this case is $\frac{1}{6}$ of 6000, which works out to be 1000.

ii. By the central limit theorem, as the size of the sample approaches infinity, the random variable will become normally distributed, and since the mean of the coin toss should approach $frac12$, the shift on the left side allows the distribution to be centered at 0.

9 Problem 9

Solution: Solution to problem 9 See following page for solutions.

10 Problem 10

Solution: Solution to problem 10 See following page for code and graphs.

11 Problem 11

Solution: Solution to problem 11

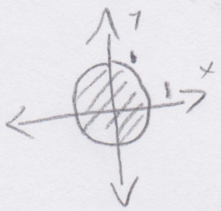
$$\begin{bmatrix} 0 \\ 0.89442719 \end{bmatrix}$$

9 Linear Algebra

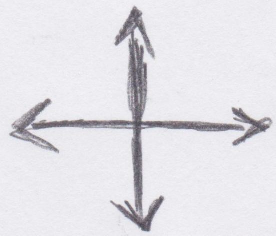
(a) **Vector Norms**
Draw the regions corresponding to vectors $\boldsymbol{x} \in \mathbb{R}^2$ with following norms:

i. $||\boldsymbol{x}||_2 \leq 1$ (Recall $||\boldsymbol{x}||_2 = \sqrt{\sum_i x_i^2}$.)
ii. $||\boldsymbol{x}||_0 \leq 1$ (Recall $||\boldsymbol{x}||_0 = \sum_{i:x_i \neq 0} 1$.)
iii. $||\boldsymbol{x}||_1 \leq 1$ (Recall $||\boldsymbol{x}||_1 = \sum_i |x_i|$.)
iv. $||\boldsymbol{x}||_\infty \leq 1$ (Recall $||\boldsymbol{x}||_\infty = \max_i x_i$.)

i.

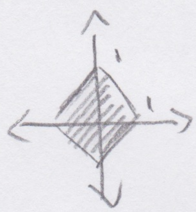


ii.

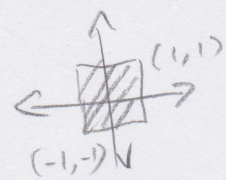


(the axes)

iii.



iv.



(b) **Matrix Decompositions and Rank**

eigenvalue λ and eigenvector x are such that
$Ax = \lambda x \quad , x \neq 0$

i. Give the definition of the eigenvalues and the eigenvectors of a square matrix.
ii. Find the eigenvalues and eigenvectors of

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

eigenvalues: $3, 1$
eigenvectors: $\begin{bmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix}, \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$

iii. For any positive integer k, show that the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$, the k^{th} powers of the eigenvalues of matrix $\boldsymbol{A}$, and that each eigenvector of $\boldsymbol{A}$ is still an eigenvector of $\boldsymbol{A}^k$.

$A^2x = A(Ax) = A(\lambda x) = \lambda(Ax) = \lambda^2 x$

We can unpack $A^k x = \lambda^k x$ in such a way to show that this equality holds true, and that ~~[illegible]~~ $\lambda_1^k, \lambda_2^k, \ldots$ are the eigenvalues and $x_1, x_2 \ldots$ are the eigenvectors of A^k.

(c) **Vector and Matrix Calculus**

Consider the vectors $\boldsymbol{x}$ and $\boldsymbol{a}$ and the symmetric matrix $\boldsymbol{A}$.

i. What is the first derivative of $\boldsymbol{a}^T\boldsymbol{x}$ with respect to $\boldsymbol{x}$?

ii. What is the first derivative of $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ with respect to $\boldsymbol{x}$? What is the second derivative?

i. Let $a^T x$ be a function $f(x)$

first derivative: $\nabla_x f(x)$

ii.

(d) **Geometry**

i. Show that the vector $\boldsymbol{w}$ is orthogonal to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$. (Hint: Consider two points $\boldsymbol{x}_1, \boldsymbol{x}_2$ that lie on the line. What is the inner product $\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$?)

ii. Argue that the distance from the origin to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ is $\frac{b}{||\boldsymbol{w}||_2}$.

i. $x_2 = ax_1 + b$

$ax_1 - x_2 + b = 0$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad w = \begin{bmatrix} a \\ -1 \end{bmatrix}$

$\begin{bmatrix} a & -1 \end{bmatrix} \begin{bmatrix} x_1 - y_2 \end{bmatrix} = \begin{bmatrix} ax_1 - ay_2 - x_1 + y_2 \end{bmatrix}$

true

ii. $wx + c = 0$

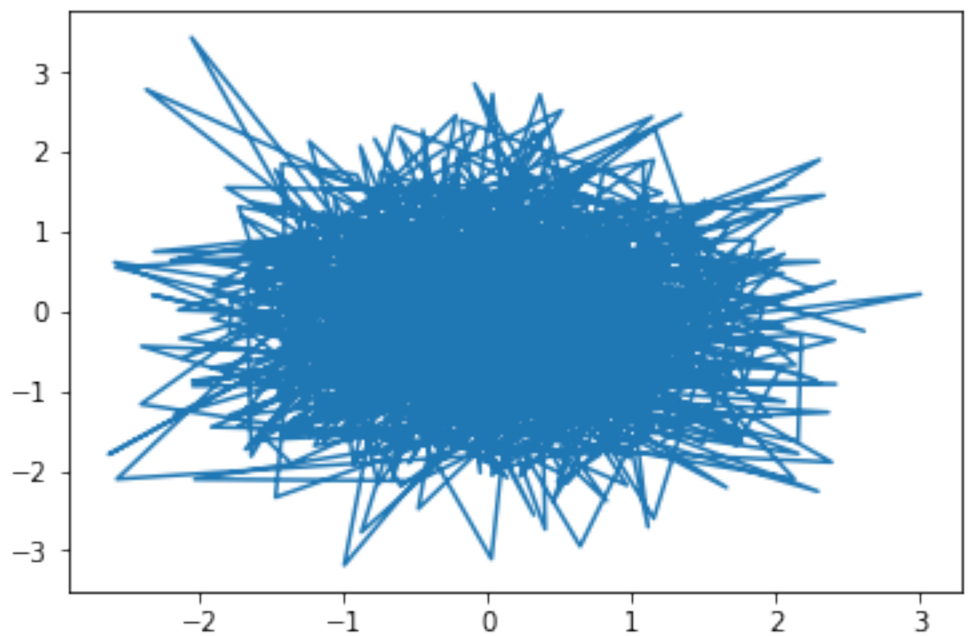
pset0

January 19, 2017

```
In [1]: %matplotlib inline
        import numpy as np
        import matplotlib.pyplot as plt
```

```
In [3]: mean_0 = (0, 0)
        cov_identity = [[1,0],[0,1]]
        x1, x2 = np.random.multivariate_normal(mean_0, cov_identity, 1000).T
        plt.plot(x1, x2)
```

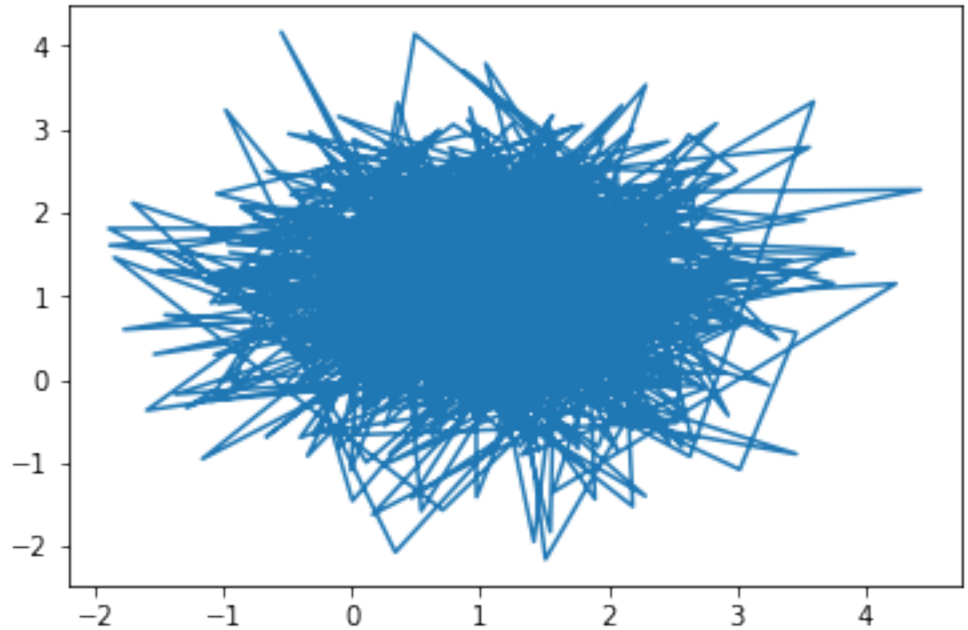
```
Out[3]: [<matplotlib.lines.Line2D at 0x1107d9c10>]
```




```
In [4]: mean_1 = (1, 1)
```

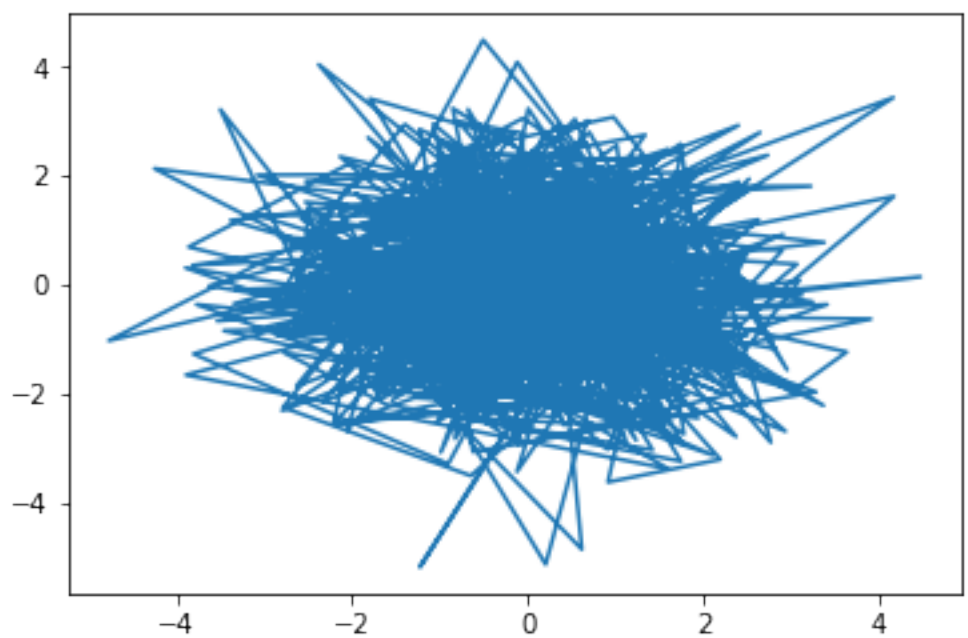
```
In [5]: x1, x2 = np.random.multivariate_normal(mean_1, cov_identity, 1000).T
        plt.plot(x1, x2)
```

Out[5]: [<matplotlib.lines.Line2D at 0x110c02090>]



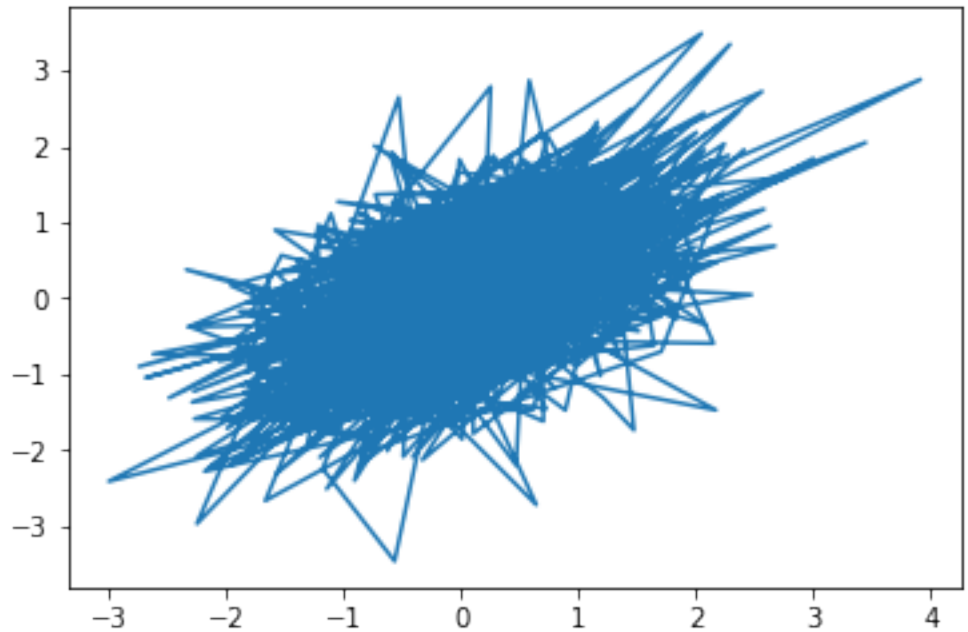

```
In [6]: cov_doubled = [[2,0],[0,2]]
        x1, x2 = np.random.multivariate_normal(mean_0, cov_doubled, 1000).T
        plt.plot(x1, x2)
```

Out[6]: [<matplotlib.lines.Line2D at 0x110e7d390>]



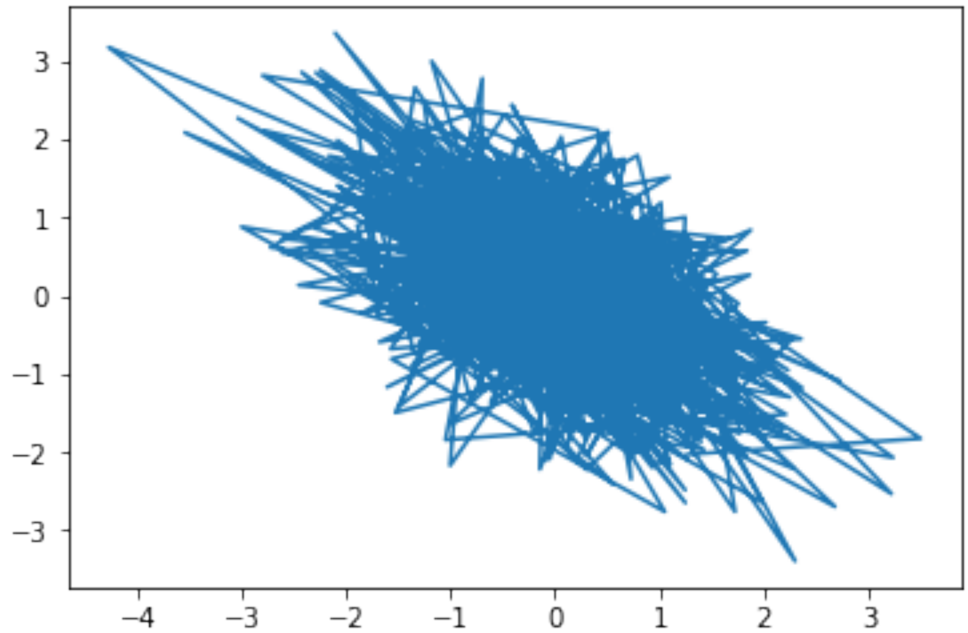
```
In [7]: cov_d = [[1, 0.5],[0.5, 1]]
        x1, x2 = np.random.multivariate_normal(mean_0, cov_d, 1000).T
        plt.plot(x1, x2)
```

```
Out[7]: [<matplotlib.lines.Line2D at 0x1110afd50>]
```




```
In [8]: cov_e = [[1, -0.5], [-0.5, 1]]
        x1, x2 = np.random.multivariate_normal(mean_0, cov_e, 1000).T
        plt.plot(x1, x2)
```

```
Out[8]: [<matplotlib.lines.Line2D at 0x11114c450>]
```




```
In [9]: prob_11 = [[1, 0], [1, 3]]
        eig_vals, eig_vects = np.linalg.eig(prob_11)
        eig_vals; eig_vects

Out[9]: array([[ 0.        ,  0.89442719],
               [ 1.        , -0.4472136 ]])

In [10]: eig_vals

Out[10]: array([ 3.,  1.])

In [11]: eig_vects

Out[11]: array([[ 0.        ,  0.89442719],
                [ 1.        , -0.4472136 ]])

In [ ]:
```

12 Problem 12

(a) Problem 12a **Solution:** Solution to problem 12a

Book crossing dataset.

(b) Problem 12b **Solution:** Solution to problem 12b

http://www2.informatik.uni-freiburg.de/ cziegler/BX/

(c) Problem 12c **Solution:** Solution to problem 12c

Although the dataset is given as a couple tables, they can be joined to have each rating as a sample. The features for each sample include the rating, the geographic location of the rater, the age of the rater, and the book itself, which could allow us to predict what kinds of books a certain person of a certain age/geographic location would enjoy.

(d) Problem 12d **Solution:** Solution to problem 12d

1,149,780 ratings.

(e) Problem 12e **Solution:** Solution to problem 12e

There are 4 features for each sample.