Assignment 3: Write Up

By Ravi Patel
PID: A11850926

NOTE: All code for this assignment will be linked separately from this write-up.

1. Entropy and Mutual Information

1.1.

$$H(X) = -\sum_{i=1}^{4} P(X=i)log(P(X=i)) =$$

$$-P(X=1)log(P(X=1)) - P(X=2)log(P(X=2)) -$$
$$P(X=3)log(P(X=3)) - P(X=4)log(P(X=4))$$
$$= -0.25 * log(0.25) - 0.30 * log(0.30) - 0.20 * log(0.20) - 0.25 * log(0.25)$$
$$= 0.598$$

1.2

$$H(Y) = -\sum_{i=1}^{4} = P(Y=i)log(P(Y=i))$$
$$-P(Y=1)log(P(Y=1)) - P(Y=2)log(P(Y=2)) -$$
$$P(Y=3)log(P(Y=3)) - P(Y=4)log(P(Y=4))$$
$$= -0.3 * log(0.3) - 0.15 * log(0.15) - 0.35 * log(0.35) - 0.2 * log(0.2)$$
$$= 0.58$$

1.3

$$H(X,Y) = -\sum_{i=1}^{4}\sum_{j=1}^{4} P(X=i, Y=j)log(P(X=i, Y=j))$$

$$= -\Big[0.15 * log(0.15) + (4)(0.03) * log(0.03) + (4)(0.05) * log(0.05) + 0.07 *$$
$$log(0.07) + (4)(0.02) * log(0.02) + 0.2 * log(0.2) + (2)(0.1) * log(0.1)\Big]$$
$$= 1.123$$

Then

$$H(X|Y) = H(X,Y) - H(Y) = 1.123 - 0.58 = 0.543$$

1.4

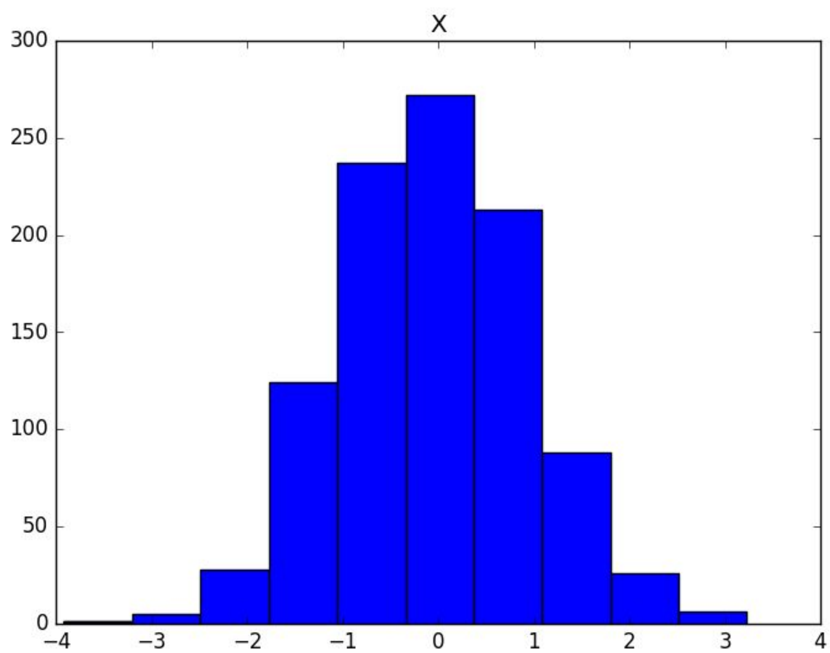
$$H(Y|X) = H(X,Y) - H(X) = 1.123 - 0.598 = 0.543$$

1.5

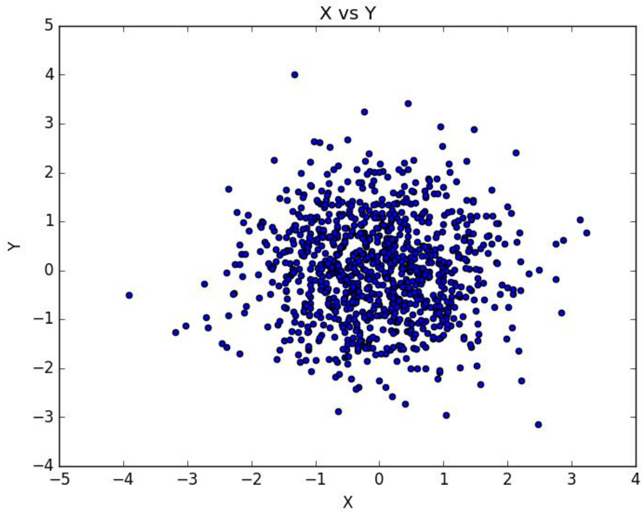
$$I(X;Y) = H(X) - H(X|Y) = 0.598 - 0.543 = 0.055$$

2. Convex or non-convex
    2.1. Convex
    2.2. Non-convex
    2.3. Non-convex
    2.4. Convex
    2.5. Non-convex
    2.6. Convex

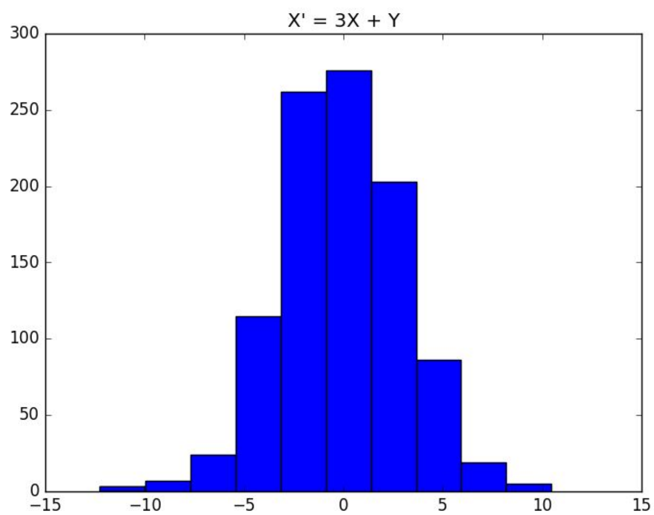
3. Gaussian Distribution: from this little experiment, one can see some useful properties holding true for Normal Distribution.



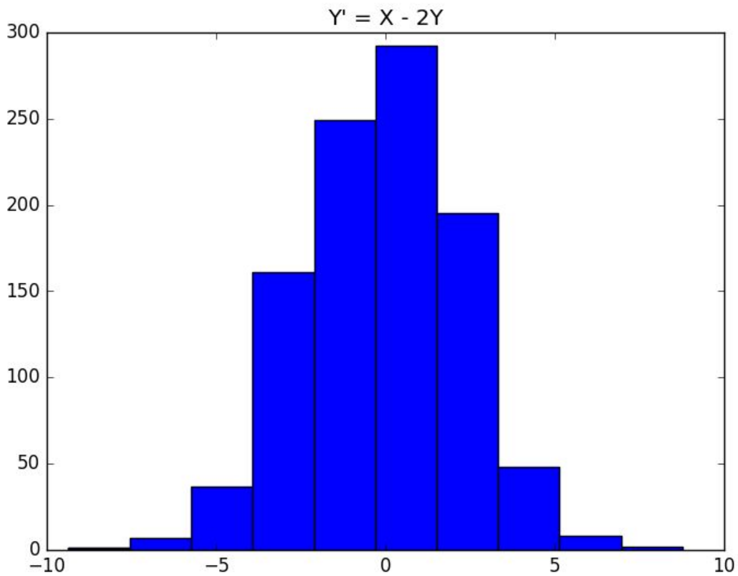
This histogram above is formed from randomly generated 1000 points from a normal distribution with mean 0 and standard deviation 1. Notice that most of the points are clustered around zero and is roughly spread by one unit to the left and right moving further away from zero. This is perfect indication of what normal distribution should be, as it is meant to be Gaussian distributed.



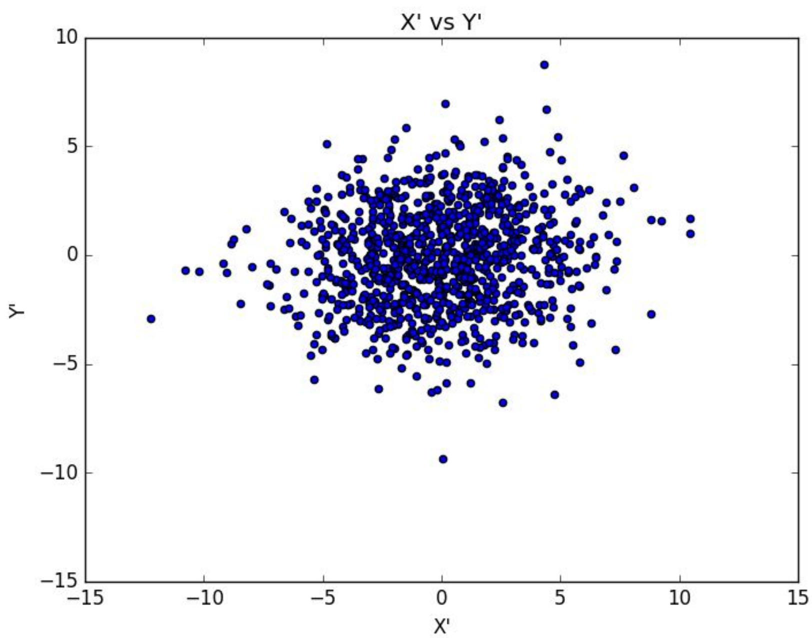
The scatter plot above is generated from two normally distributed random variables, both with mean 0 and variance 1. Notice a good portion of points are clustered towards the center. Also, as one moves away from the center in any direction by one unit, the occurrence of points gradually declines - a typical behavior of Gaussian Distribution with mean 0 and standard deviation 1.



One useful property of normal distribution is that mean and variance is closed under addition. Its being seen in the above histogram. Multiplying X by 3 would also multiply the mean by 3 and adding that to Y would increase the mean by the corresponding amount. Similar computation is done on variance. In the end, mean is 0 and variance is 10. That can be seen in this figure. Most of the points are clustered around zero and moving away to left or right by approximately 3 units would gradually reduce the occurrence of points - just the behavior expected from normal distribution.

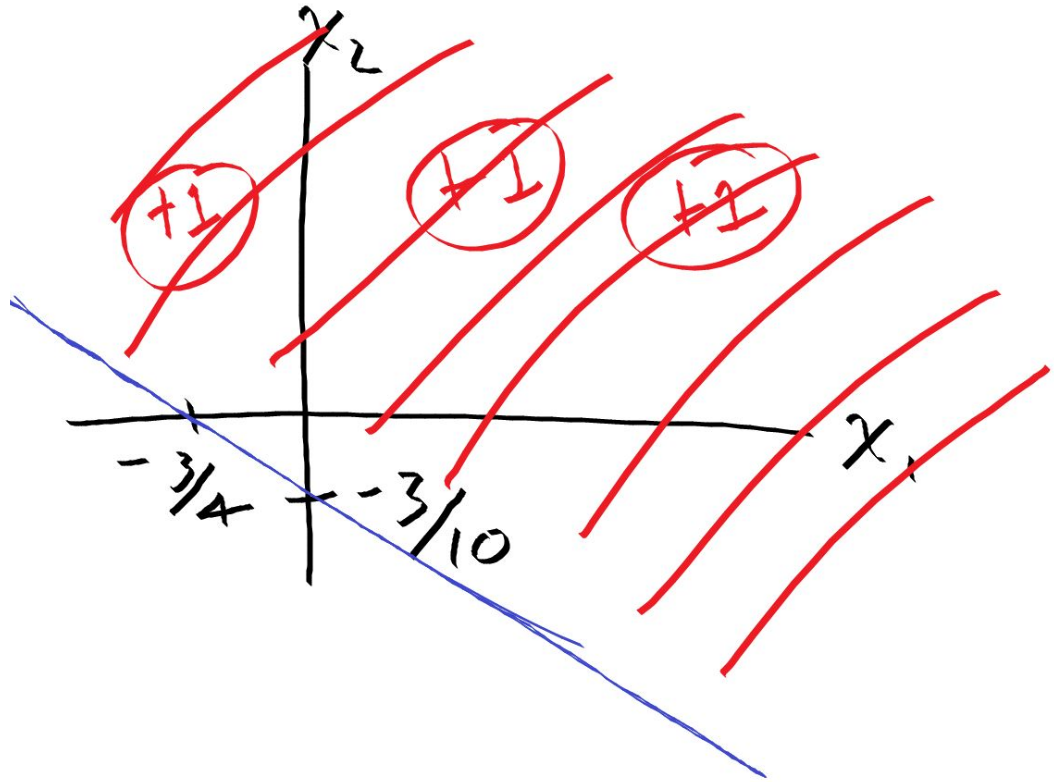


A similar explanation as above.

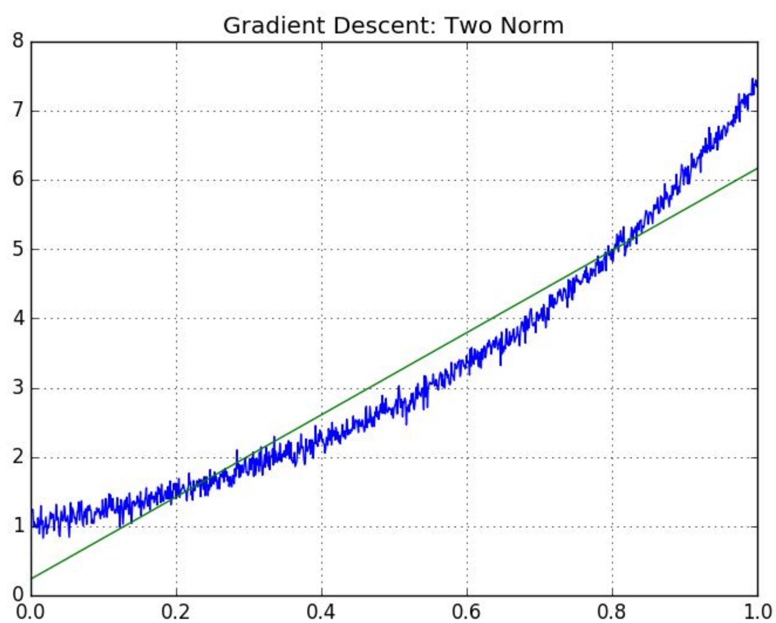


Similar explanation as from the scatter plot from the previous point.

4.



5. L2-Norm Gradient Descent



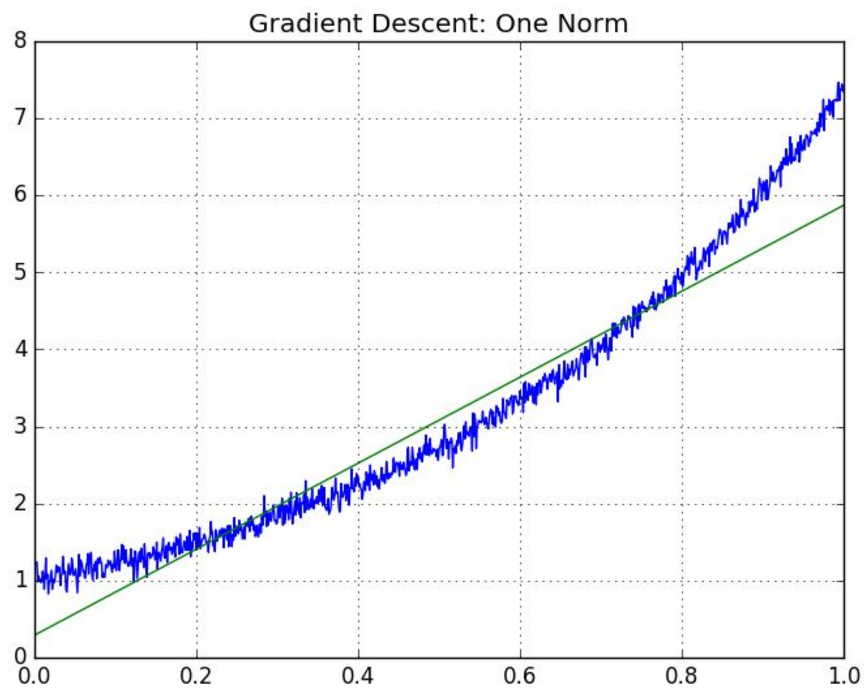
The dimension of both W and $\frac{dg(W)}{dW}$ is 2 *x* 1

Assignment 2: W = (0.207 ; 5.981)

L2-Norm Gradient Descent: W = (0.234 ; 5.931)

Comparing the two results, gradient descent approximation is very close to the exact closed form solution. The value of W computed via gradient descent, two norm, seems to makes sense. Since gradient descent is an *approximation*, it will not provide the exact same solution as the closed form solution; however, the approximation is close to the exact solution.

6. L1-Norm Gradient Descent



L1-Norm Gradient Descent: W = (0.28 ; 5.58)

Comparing the one norm gradient descent solution to that from the previous problem and closed formed solution, this is also relatively close to the exact solution. This value of W computed via gradient descent, one norm, seems to makes sense. Since one norm is less susceptible to outliers, the solution to one norm should generate regression line closer to center cluster and it does, indicated by a small shift downward of the regression line from the previous problem.