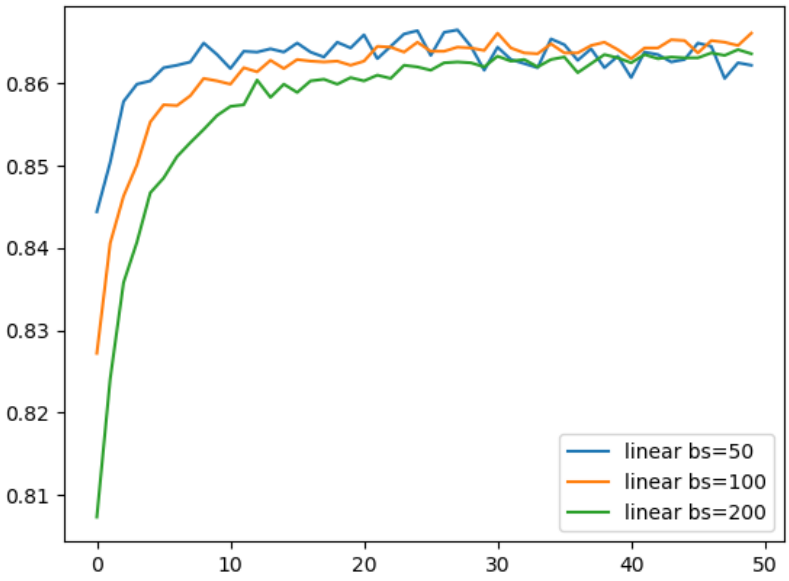
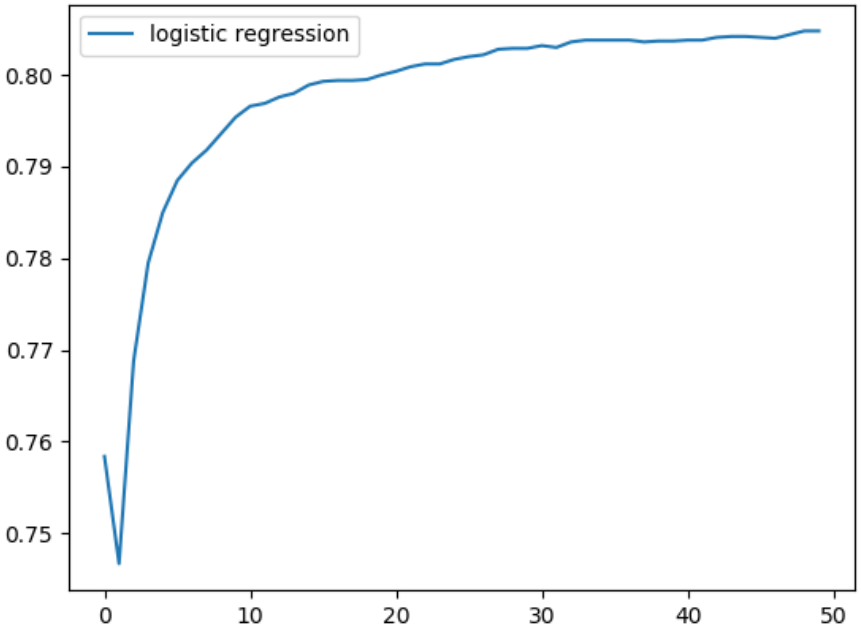
3 (a)



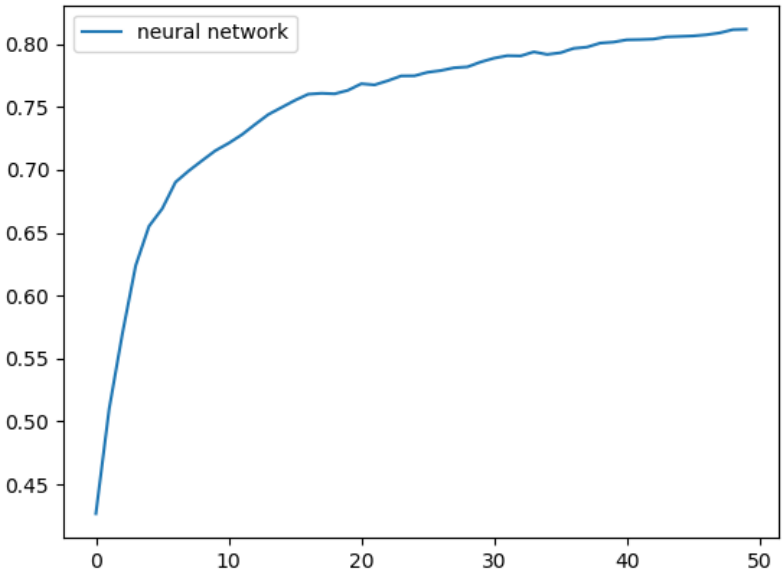
running time for batch size of 50: 63.273s
running time for batch size of 100: 47.256s
running time for batch size of 200: 42.022s

Using small batch size, the potential benefits are faster convergence, but the drawbacks are longer calculation time

(b)



(c)



(e)
1.
learning_rate = 0.01
training_epochs = 100
batch_size = 100

Epoch: 0015 cost= 57.841941198 |w-w_true|^2 = 0.017615196
TLS through SVD error: |w-w_true|^2 = 0.0003320206394634698

2. change learning rate
learning_rate = 0.05
training_epochs = 100
batch_size = 100

Epoch: 0004 cost= 56.934029833 |w-w_true|^2 = 0.051046396
TLS through SVD error: |w-w_true|^2 = 0.0003320206394634698

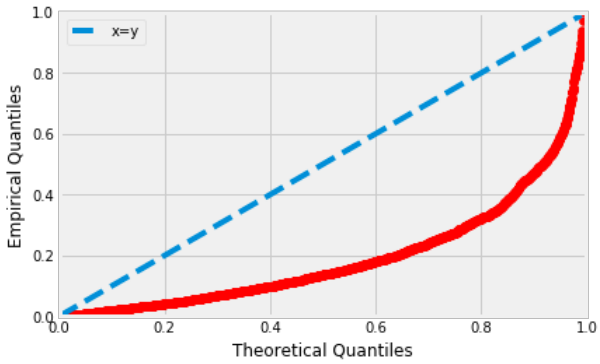
3. change batch_size
learning_rate = 0.01
training_epochs = 100
batch_size = 500

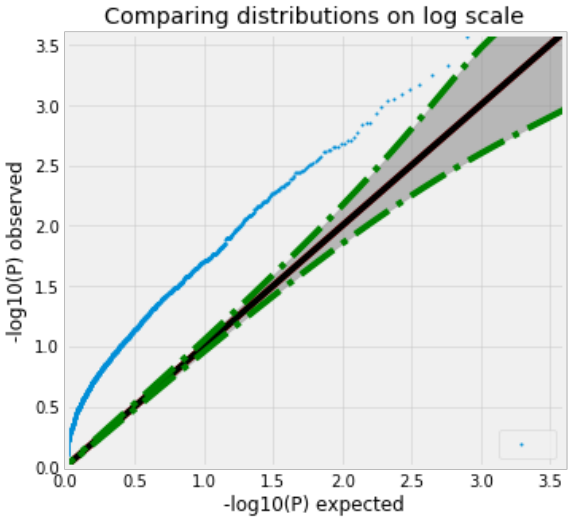
Epoch: 0086 cost= 248.381114960 |w-w_true|^2 = 0.014546557
TLS through SVD error: |w-w_true|^2 = 0.0003320206394634698

the solution of SGD is not sensitive to its hyper-parameters in this problem

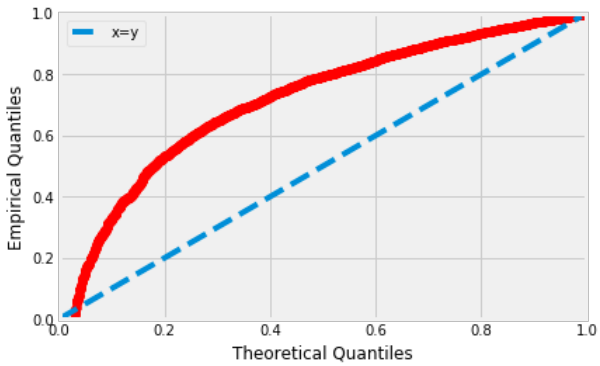
4 (a)

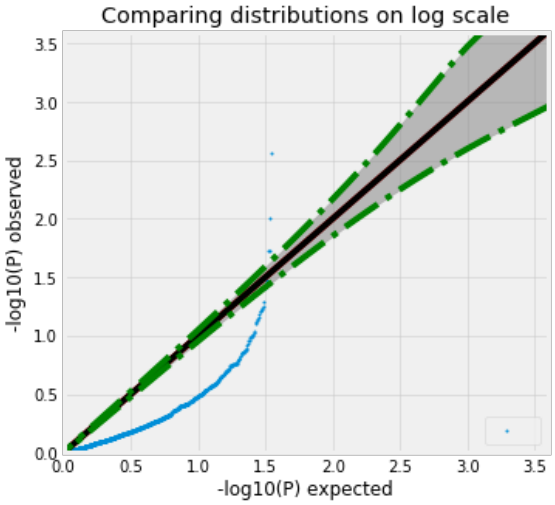
Skewleft



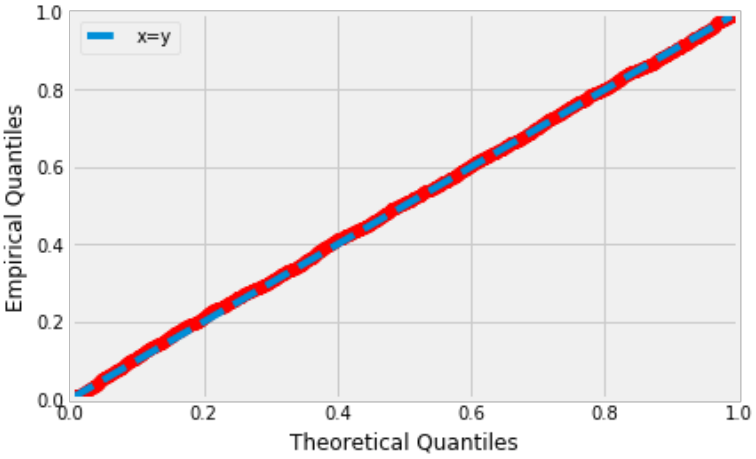


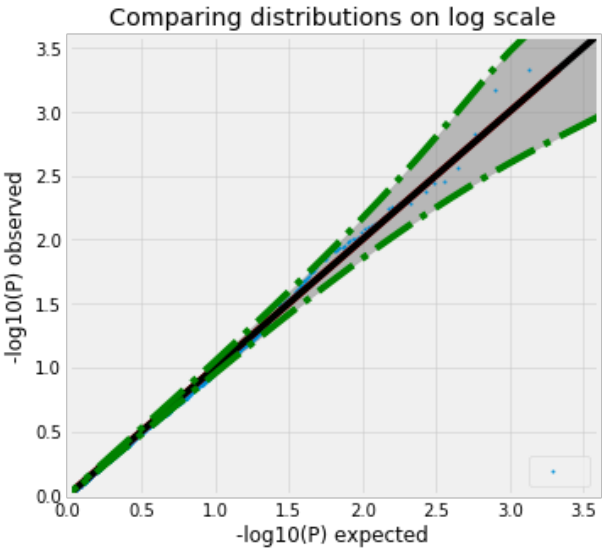
Skew right





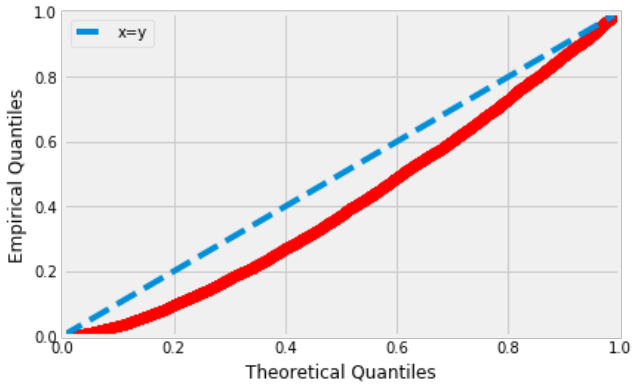
uniform

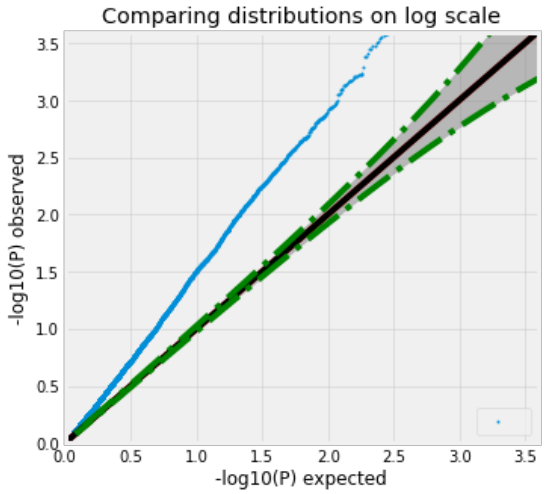




I found that for uniform distribution, then the empirical quantiles lies on top of theoretical quantiles, meaning that. For Left-skewed distribution or right-skewed distortion, the empirical quantiles are different with theoretical quantiles.
We should see the observed line gets more and more closer to theoretical line if our empirical distribution looks more and more similar to a Unif[0, 1] .

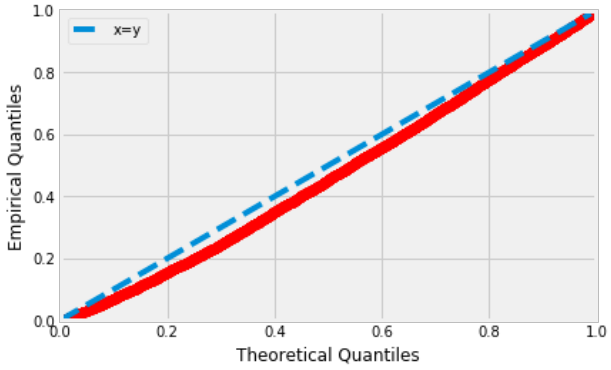
(b)

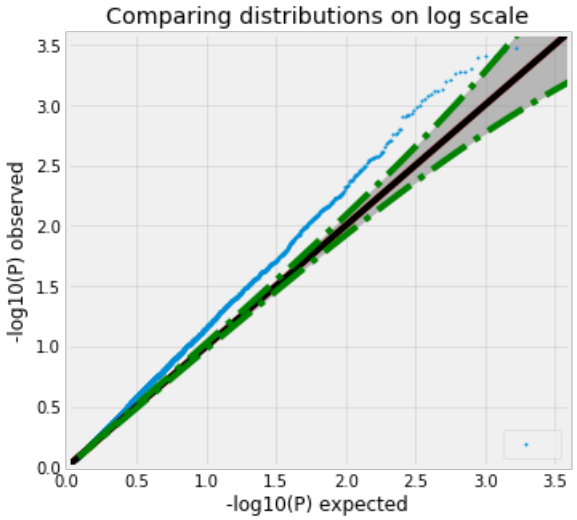




From the plot, we see the difference between empirical and theoretical quantiles, so linear regression is not suitable for our problem.

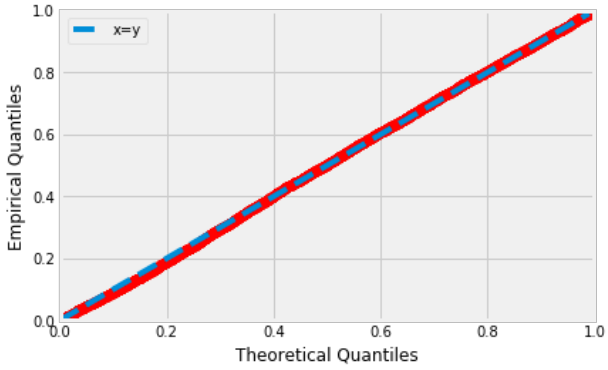
(c)
Using all genetic features will be too many free parameters, it's computationally expensive. More importantly, we'll have more variants and data, so we will overfit our data.

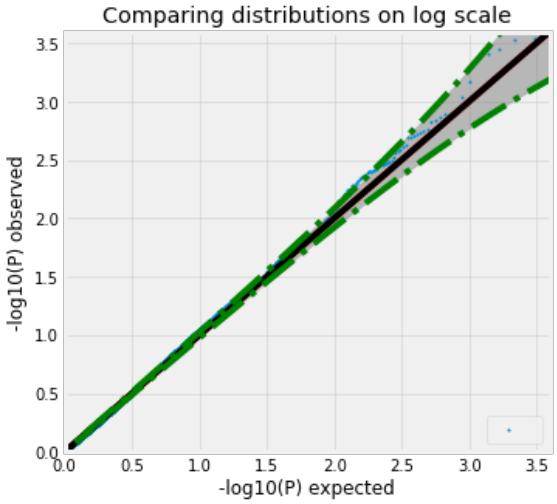




This plot is the quantiles agrees better, so PCA is a better model than pure linear regression.

(d)





From the plot, we can see this model is the best among the three models.