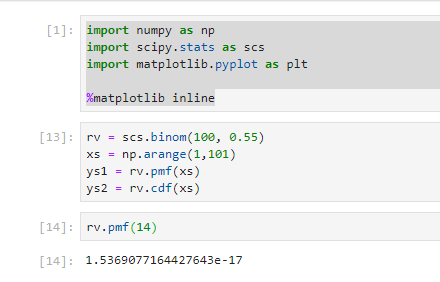
**Quiz**

1. A vector is given, [2,3,9]. What's the norm?

*9.695*

1. I have a biased coin. P(T) = 0.45. What's the probability, if I flip the coin 100 times, that I will get exactly 14 heads?



1. What is the rank of a matrix?

*Number of independent rows or columns*

1. Why is rank important for Linear Regression?

*If the columns or rows are dependent then multicollinearity exists between the features and it will increases the standard errors of the coefficients will make the model unstable.*

1. What is the purpose of the sigmoid function in a Logistic Regression model?

*To calculate the probabilities.*

1. What is bias vs variance?

*Bias is the difference between the average prediction of our model and the correct value which we are trying to predict. Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.*

*Variance is the variability of model prediction for a given data point or a value which tells us spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn’t seen before. As a result, such models perform very well on training data but has high error rates on test data.*

1. What is a hyperparameter? What part of the dataset (train, validation or test) is responsible for determining this value?

*Model hyperparameters, on the other hand, are common for similar models and cannot be learnt during training but are set beforehand. A typical set of hyperparameters for NN include the number and size of the hidden layers, weight initialization scheme, learning rate and its decay, dropout and gradient clipping threshold, etc.*

*Training dataset is responsible for determine this value.*

1. Model selection via cross validation. What part of the dataset (train, validation or test) is responsible for choosing the best model?

*Training dataset*

1. What is parametric vs non-parametric model?

*A parametric model captures all its information about the data within its parameters. All you need to know for predicting a future data value from the current state of the model is just its parameters. For example, in case of a linear regression with one variable, you have two parameters (the coefficient and the intercept). Knowing these two parameters will enable you to predict a new value.*

*Example: Linear and logistic regression  
  
On the other hand, a non-parametric model can capture more subtle aspects of the data. It allows more information to pass from the current set of data that is attached to the model at the current state, to be able to predict any future data. The parameters are usually said to be infinite in dimensions and so can express the characteristics in the data much better than parametric models. It has more degrees of freedom and is more flexible.*

*Examples: Random forest, decision tress, ada boosting and NN*

1. What models do you need to scale the data prior to fitting?

*Linear, Logistic, KNN and NN*

1. What is entropy? Which model uses this concept?

Entropy controls how a Decision Tree decides to **split**the data. It actually effects how a **Decision Tree** draws its boundaries. Decision trees use entropy.

1. How is a random forest generated?

*Random forest is generated using bagging procedure. Each tree is built on a sample of objects drawn with replacement from the original set; thus each tree has some objects that it hasn't seen, which is what makes the whole ensemble more heterogeneous and thus better in generalizing.*

1. How do you determine the correct number of clusters when using KMeans?

*By using elbow curve. The location of a bend in the plot is generally considered as an indicator of the****appropriate number of clusters****.*

1. What is a dendrogram?

*A****dendrogram****is a diagram that shows the hierarchical relationship between objects.*

1. What is linkage?

*The process of selecting two individual elements/clusters in hierarchical clustering to form a new cluster is called linkage. Various types of linkage are: single, complete, average, centroid*

1. How does a recommender model work?

*Recommender system looks at user’s ratings (in training data) and tries to predict rating of an item that user in question has not bought before. Based on predicted ratings, items are recommended to the user in question*

1. How can you reduce the number of features in a model?

*Using PCA or Lasso*

1. What is a good use of PCA?

*To reduce the dimensionality*

1. In general, how do neural networks "learn"?

*Neural networks break the learning process into multiple layers. At each layer, it*

*takes input and activates a neuron by taking weighted approach. In the process it*

*discovers latent features. At the end layer, prediction is done by taking weighted*

*approach of previous layer. This is using forward and backward propagation to*

*improve prediction*

1. What is your favorite model? Why?

*Random forest, because in this model multiple samples of your training data are taken then models are constructed for each data sample. When you need to make a prediction for new data, each model makes a prediction and the predictions are averaged to give a better estimate of the true output value*.

*Reasons:*

1. *It can solve both regression and classification problems.*
2. *In most of the assignments I observed that the prediction was much better using random forest compared to other models.*