Exam 2 Results for Dingyu Wang

Score for this quiz: **90** out of 100
Submitted Dec 13 at 2pm
This attempt took 88 minutes.

Question 1

0 / 20 pts

For each of the below statements, agree or disagree and explain your reasoning:

- Hierarchical clustering can't handle big data well but K Means clustering can. This is because the time complexity of K Means is linear i.e. O(n) while that of hierarchical clustering is quadratic i.e. $O(n^2)$.
- In K Means clustering, since we start with random choice of clusters, the results produced by running the algorithm multiple times might differ. The same is true in Hierarchical clustering since the order of the data can be random.
- K Means clustering requires prior knowledge of K i.e. no. of clusters you want to divide your data into. But, you can stop at whatever number of clusters you find appropriate in hierarchical clustering by interpreting the dendrogram.

Your Answer:

For the first point, I agree. K mean is linear. The hierarchical clustering is quadratic. Single linking is O(N^2) and for others like complete linkage its O(N^2logN), both of which are worse then O(N). Hierarchical is slower as it has to do the merging or splitting each step.

For the second point, I agree that the K-means clustering it's sensitive to the centroid starting locations and can produce different results, we saw this in the video lectures. However, with hierarchical clustering we should get the same results each time.

For the third point, I agree. K means does require prior knowledge to finding K. typically we use the elbow method to find a suitable K. And yes, hierarchical clustering you do have the ability to stop at the desired number of clusters by looking at the dendrogram. You don't need prior knowledge on K for hierarchical.

Question 2

0 / 20 pts

For feed forward neural networks, explain what can be represented. Explain how weights are updated in the back propagation step.

Your Answer:

Feed forward neural networks can represent things like the XOR that a perceptron cannot represent. Also can represent other things with complex boundaries. A feed forward NN with 2 or more layers can represent any function. Weights are updated in the backpropagation step through gradient descent method, taking gradient of the network weights.

Question 3

0 / 20 pts

When are CNN useful? In a CNN, how are the number of connections reduced in comparison to a feed forward NN? Explain pooling layer and convolution layer.

Your Answer:

CNN are useful like in image processing. It can automatically detect features from images (like the bird beaks in the class example),

regardless of the shifts and scaling of the image. The number of connections are reduced in comparison to feed forward NN because of sharing of weights on edges. The pooling layer is when it does a subsample of the image. It makes picture smaller. There's less parameters on the image, so it could help with sensitivity of shifting and scaling. A convolutional layer has one or many filters than run across the image array to extract key features from the image. Like in the homework we can have horizontal, vertical, diagonal filters.

Question 4

0 / 20 pts

What is the difference between bias and variance? How do these relate to underfitting and overfitting? How can they help you select models? How are they used to tune hyperparameters?

Your Answer:

Bias is how good the model fits the observed data, whereas variance is how robust the model is to a particular training set. High bias models have a tendency to underfit and high variance models have a tendency to overfit. There's a tradeoff between bias and variance. We want to select model with ideally low variance and low bias. There is a point where variance and bias^2 cross on graph and that's the models we are looking for. We can do a search of all the hyperparameters and see which ones give the lowest bias and variance. That's how we tune it.

Question 5

0 / 20 pts

When are RNN useful? What is the shortcoming of a simple RNN and how is that solved using LSTM?

Your Answer:

RNN are used for time dependent and sequential data. RNN have an internal state variable. RNN takes too long to converge or could never converge. Is it also simplistic and doesn’t really model how real human brains. LSTM on other hand has a forgetting property that it could get rid of some parts of the internal state. LSTM also can have features like ignoring input, deciding on output, etc.

Quiz Score: **90** out of 100

This quiz score has been manually adjusted by +90.0 points.