**Solutions**

**Ans1:**

**Activation Function:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Ans2:**

The prior probability of an event is the probability of the event computed before the collection of new data. One begins with a prior probability of an event and revises it in the light of new data.

Conditional probability is a measure of probability to an event that is occurring, given other event has already occurred.

Also referred as conditional models. This is a class of logistical models used for classification or regression. It should distinguish decision boundary through observed data.

A generative model describes how a dataset is generated, in terms of a probabilistic model. By sampling this model, we generate data.



Beam Size

More words would be traversed and chosen. Therefore, a higher probability to get better results.

More computational resources and more memory are required.

**Ans3:**

1. **Trigram model of language modelling:** A trigram model restricts the conditional information to the previous two words. Using this method, the conditional distribution can calculate a certain word combination frequency based on the previous two words.
2. **Procedure of 5-fold cross-validation:** Dataset is split into 5 sets. Each one set is taken as test set by turn while other four are training sets. Eventually, we’ll have 5 accuracies and the average is the accuracy of 5-fold cross-validation.

**Pros:** Avoiding the randomness and bias by training and testing all the data.

1. **Bagging:** Bootstraps the training set, estimates many copies of a model on the resulting samples and then averages their predictions.

**Boosting:** Sequentially reweights the training samples forcing the model to attend the training examples with higher loss.

**Stacking:** Used a separate validation set to train a meta-model that combines predictions of multiple models.