READING ASSIGNMENT -dxm172530

Chapter 1:

1. What are the phases in the ML lifecycle? What type of evaluation is suitable for each phase?

* The first phase is prototyping where the focus is to find the best model :

Offline evaluation is suited for this phase .

* The second phase is deployment into production and testing :

Online evaluation is suited for this phase .

1. Why is the process of evaluation complicated?

The process of evaluation is complicates since

* Offline and online evaluation use different metrics from one another . Offline evaluation may use metrics like accuracy while online evaluation may focus on business metrics
* Models may have the assumption that data distribution is stationary while actually in practice data distribution changes over time . This leads to distribution drift .

3. What is the difference between *model parameter* and *hyperparameter*?

Give examples of each for different models that you have learned. For example, in case of

decision tree, what would be a *model parameter* and what would be a *hyperparameter*?

* Model parameter :

It is a property of training data and can be learnt by the algorithm .

* Hyper parameter :

It is not learned by the training algorithm but rather needs to be tuned .

For example , in a SVM – the model parameter is weights of variables and hyper parameter can be a kernel and slack variables .

Decision Tree :

Model parameters - Split points in the tree .

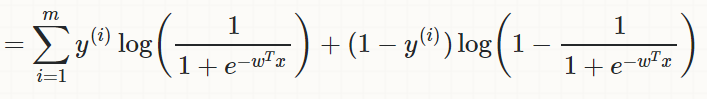
Hyper parameters - Number of attributes or depth of a decision tree

Chapter 2:

1. What would be the equation for log-loss metric for a binary class dataset when using Logistic

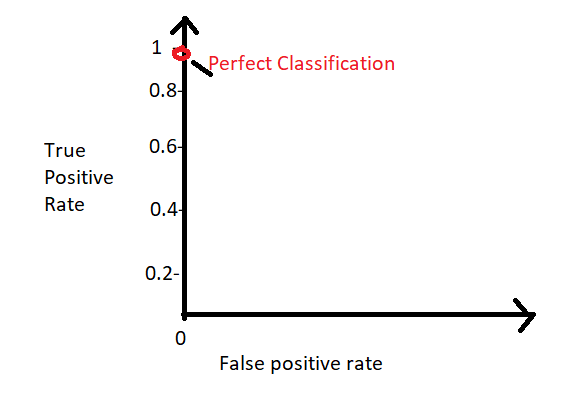
Regression as the model? Write the equation in the most simplified format.

Ans : log‐loss = − 1/ N ∑Ni=1 y(i) log P(y=1)+ (1 − y(i) ) log P(y=0)

 for logistic regression .

https://en.wikipedia.org/wiki/Logistic\_regression

2.Read the section on ROC and draw the curve for the perfect classifier i.e. one that makes no mistakes.



3. Understand the concepts of precision, recall, and F1 score. Is it possible for precision and recall both to go up at the same time i.e. do they have a positive or negative correlation? Explain.

The given formula is :

F1 = 2 \* precision\*recall / precision + recall

This clearly states that F1 score will be small only if one of the two : i.e precision or recall is low .Therefore precision and recall cannot both go up at the same time

4.What are some of the things you should be cautious about when choosing evaluation functions, and when analyzing the distribution of data.

* Know the difference between training metrics and evaluation metrics , i.e the model should not be asked to work on a task it was not assigned to .
* Lookout for skewed data : imbalanced classes , outliers and rare data .

Chapter 3:

1. Why is model selection and hyperparameter tuning done using results of validation dataset and not training dataset?

-> The model training process receives training data and produces a model, which is evaluated on validation data. The results from validation are passed back to the hyperparameter tuner, which tweaks some knobs and trains the model again.

-> Model selection and hyperparameter tuning is done on validation dataset to ensure fairness since the model is to be validated on data which it hasn’t seen before.

2.If you are given just one dataset, what are the 3 ways in which you can obtain validation dataset(s) from the given dataset? Explain each and give advantages and disadvantages of each.

The 3 ways to obtain validation datasets from the given data set is : Hold-Out Validation, Cross-Validation and Bootstrapping .

1. Hold-Out Validation :

* Assuming that all data points are independently and identically distributed, we simply randomly hold out part of the data for validation.
* Model training is done on the larger portion of the data and validation metrics is evaluated on the smaller hold-out set.

Advantages :

* Simple to program and faster to run it .

Disadvantages :

* It is less powerful statistically.
* The validation results are derived from a very small subset of the data, hence its estimate of the error is less reliable.

1. Cross Validation :

K-fold cross validation : we first divide the training dataset into k folds .For a given hyperparameter setting, each of the k folds takes turns being the hold-out validation set; a model is trained on the rest of the k – 1 folds and measured on the held-out fold.

Advantages :

It is useful when the training dataset is so small that one can’t afford to hold out part of the data just for validation purposes.

Disadvantages :

Not much of use for a larger dataset .

1. Bootstrap :

Bootstrap is a resampling technique. It generates multiple datasets by sampling from a single, original dataset. Each of the “new” data‐ sets can be used to estimate a quantity of interest.

Advantages :

We can also calculate variance or a confidence interval for the estimate.

Disadvantages :

The same data point may be present more than once in the bootstrapped set .

3. We live in an era of Big Data. Why don’t we just get more data rather than obtaining

validation data from given data?

* Even though we live in an era of Big Data , the techniques like Bootstrapping and Cross validation resample the data to produce multiple data sets ,and hence are still relevant for evaluation mechanisms .
* We need to be frugal with data .

4. Why should training, and evaluation data never be mixed?

If information from the validation data or test data leaks into the training procedure, it would lead to a bad estimate of generalization error.

Chapter 4:

1. What is the role of a *hyperparameter* in a model? When are model parameters learned? Do you think a model is more strongly influenced by *model parameters* or *hyperparameters*? Explain you reasoning?

Note: There is no right or wrong answer, but you should explain your thinking.

-> Hyperparameter : A hyperparameter controls the capacity and flexibility of the model.

-> Model parameter : Model parameters are learned during the training phase .

-> Hyperparameter influences a model more strongly than model parameter , because it can be tuned ,set by some heuristics and they also help in estimation of model parameters .

For Eg : in a decision tree , the depth of a decision tree or the number of leaves plays a major role in data classification .

2. What are some techniques for hyperparameter tuning? Explain each briefly and in your own words.

Some techniques for hyperparameter tuning are :

* Grid Search :

It implements exhaustive searching through a subset of hyperparameter space before returning a winner .

* Random Search :

It is a variation of grid search . Instead of searching over the entire grid , this only searches over a sample of points on the grid .

* Smart hyperparamter tuning techniques :

Some techniques of smart hyperparameter tuning techniques :

Derivative-free optimization : Uses heuristics to determine where to sample next.

Bayesian optimization and random forest smart tuning : Both model the response surface with another function, then sample more points based on what the model says.