HW2-695

Shangxing Jiang

1. Machine Learning: With a given data set D and specific task T, we establish a model and design an algorithm based on statistics, then use some samples in D as training data to get coefficients or classifications. Finally, the rest data can be used to test the model, measuring its performance.

2. (1) Supervised learning maps an input to an output based on example input-output pairs. Data has been classified or labeled by human (or some apps). Hypothesis - target

(2) Unsupervised learning, whose data is not classified or labeled, finds some common features from the data set and reacts based on the common features to new data. eg. Predict user’s like

(3) Online learning updates its model based on one unit of data, then go to the next step, updating with next unit of data. eg. Stochastic gradient descent

(4) Batch learning updates its model with the whole training data set, then go to next step.

(5) Model-based learning has an agent try to understand the world and create a model to represent it, trying to capture transition function and reward function.

(6) Instance-based Learning, instead of generalizing a model, compares new instances with instances seen.

3. Overfitting: The model is too specific, so that it fits the training data perfectly, but has large ERMS for testing data.

Regularization: is used to prevent the model from overfitting the training sample, which is a parameter used to minimize ERMS.

4. Cross-validation: In training data, most of them (train\_set) are used to train the model and the rest (valid\_set), relatively small part, are used to test the model. Then change the train\_set and valid\_set.

One Strategy is equally dividing the data set to several part. Sequentially select each part as valid\_set.

5. (1) Not enough data, (2) The limit of computing ability and the slow running speed of python, (3) Not easy to decide the general or specific extent of a model

6. S0 = {<o, o, o, o, o, o>} v G0 = {<?, ?, ?, ?, ?, ?>}

X1 = {Sunny, Warm, Normal, Strong, Warm, Same} +

S1 = {<Sunny, Warm, Normal, Strong, Warm, Same>} v G1 = {<?, ?, ?, ?, ?, ?>}

X2 = {Sunny, Warm, High, Strong, Warm, Same} +

S2 = {<Sunny, Warm, ?, Strong, Warm, Same>} v G2 = {<?, ?, ?, ?, ?, ?>}

X3 = {Rainy, Cold, High, Strong, Warm, Change} -

S3 = {<Sunny, Warm, ?, Strong, Warm, Same>} v G3 = { <Sunny, ?, ?, ?, ?, ?> <?, Warm, ?, ?, ?, ?> <?, ?, ?, ?, ?, Same> }

X4 = {Sunny, Warm, High, Strong, Cool, Change} +

S4 = {<Sunny, Warm, ?, Strong, ?, ?>} v G4 = {<Sunny, ?, ?, ?, ?, ?> <?, Warm, ?, ?, ?, ?> }