1 Supervise Learning

1.1 2 approaches and 2 models

Supervise learning aims to learn a model and predict output of the input. The form of the model is usually like $Y = f(X)$ or conditional probability distribution $P(Y|X)$.

And supervised learning has 2 approaches:

1. Generative approach

2. Discriminative approach

Therefore 2 corresponding models:

1. Generative model

2. Discriminative model

1.2 Generative model

We learn joint probability distribution $P(X, Y)$ from data, and compute conditional probability distribution $P(Y|X) = \frac{P(Y,X)}{P(X)}$.

This approach called generative approach is because model represents the generative relationship between fixed input X and corresponding output Y.

Common generative model:

1. Naive Bayesian

2. HMM

1.3 Discriminative model

We learn decision function $f(X)$ or conditional probability distribution $P(Y|X)$ as prediction model.

Discriminative model focuses on what would the output Y be if we fix the input X.

Common discriminative model:

1. KNN

2. Perceptron

3. Decision tree

4. Logistic regression

5. Maximum entropy model

6. SVM

7. CRF

1.4 Properties of Generative approach and Discriminative approach

Properties of generative approach:

1. Learn joint probability distribution $P(Y, X)$

2. Faster convergence than discriminative approach when data increases.

3. could be applied with hidden variables but discriminative model can't.

Properties of discriminative approach:

1. Learn conditional probability distribution $P(Y|X)$ or decision function $f(X)$ directly, higher accuracy.

2. Abstracts data, defines features, utilizes features,; therefore simplify learning problem.

1.5 Classification

2 parts:

1. Learning, use learning approach and known data.

2. Classify, use learned model to predict class of new input.

Common approaches:

1. KNN

2. Perceptron

3. Naive Baysian

4. Decision tree

5. Decision tree

6. Logisitic regression

7. SVM

8. Boosting

9. Bayesian Network

10. Neural Network

11. Winnow

1.6 Tagging

Predicting output sequence with input sequence. The model is the conditional probability distribution $P(Y^{(1)}, Y^{(2)}, ..., Y^{(n)} | X^{(1)}, X^{(2)}, ..., X^{(n)})$.

Common approaches:

1. HMM

2. CRF

1.7 Regression

Regression model is a function that represents the mapping from input variable to output variable.

The most common loss function $squared loss function$, which could be solved by least squares.