

Problem 4

In General, A Discriminative model models the decision boundary between the classes. A Generative Model explicitly models the actual distribution of each class.

A Generative Model learns the **joint probability distribution** $p(x, y)$. It predicts the conditional probability with the help of Bayes Theorem. A Discriminative model learns the **conditional probability distribution** $p(y|x)$. Both of these models were generally used in supervised learning problems.

generative models are often used when we have a notion of the underlying distribution of the data and we want to find the hidden parameters of that distribution, while discriminative models are more suitable when we only want to find the boundary that separates the data into different classes.

Comparison

1. Accuracy

Generative models are less accurate than discriminative models when the assumption of conditional independence is not satisfied. For example, in spam classification problem, let $x_1 = \text{number of times "bank" appear in the email data}$, and $x_2 = \text{number of times "account" appear in the email}$. Regardless of whether spam, these two words always appear together, i.e. $x_1 = x_2$. Learning in naive Bayes results in $p(x_1 | y) = p(x_2 | y)$, which double counts the evidence. Logistic regression doesn't have this problem because it can set either $\alpha_1 = 0$ or $\alpha_2 = 0$.

2. Missing Data

Generative models can work with missing data, and discriminative models generally can't. In generative models, we can still estimate the posterior by marginalizing over the unseen variables. However, discriminative models usually require all the features X to be observed.

3. Performance

Compared with discriminative models, generative models need less data to train. This is because generative models are more biased as they make stronger assumptions (assumption of conditional independence)

4. Application

Discriminative models are “discriminative” because they are useful but only useful for discriminating Y 's label, so they can only solve classification problems. Generative models have more applications besides classification, e.g. samplings, bayes learning, MAP inference.

Machine Learning