# On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes

Andrew Ng and Michael Jordan

Generative: build a model, then extract classifier from model.

Discriminitave: build a classifier directly.

Generative: build a model, then extract classifier from model.

Discriminitave: build a classifier directly.

Prevailing wisdom: discriminative classifiers are better.

Generative: build a model, then extract classifier from model.

Discriminitave: build a classifier directly.

Prevailing wisdom: discriminative classifiers are better.

Paper says: not always

Generative = naive bayes = $h_{gen}(x) = \sum_{i=1}^{n} \ln \frac{\hat{p}(x_i|y=1)}{\hat{p}(x_i|y=0)} + \ln \frac{\hat{p}(y=1)}{\hat{p}(y=0)}$

Discriminative = logistic regression = $h_{dis}(x) = \sum_{i=1}^{n} \theta_i x_i + \theta_0$

Observation: asymptoticallly $e(h_{dis,\infty}) \leq e(h_{gen,\infty})$

VC says:

$$e(h_{dis,m}) \leq e(h_{dis,\infty}) + O(\sqrt{\frac{n}{m} \log \frac{m}{n}})$$

$$e(h_{gen,m}) \leq e(h_{gen,\infty}) + O(\sqrt{\frac{\log n}{m}})$$