Generative and Discriminative Classifiers

Illustration

- 2 classes: bird and cat
- Generative classifier
  - Training: build a bird model and a cat model – what they look like
  - Classification: match the new animal against the bird model and against the cat model; and determine if it looks more like a bird or a cat
- Discriminative classifier
  - Training: find a decision boundary separating birds and cats
  - Classification: check which side of the decision boundary the new animal falls and assign the label accordingly

Description

- Generative classifier
  - *p(x|y)* and *p(y)* to represent the joint probability then use Bayes rule to obtain *p(y|x)*
  - Examples: Naïve Bayes, HMM (sequence model/sequence of observations)
- Discriminative classifier
  - Learn *p(y|x)* directly from data
  - Examples: Logistic Regression, Conditional Random Field (sequence model)

Parametric and Non-Parametric

- Parametric model
  - The number of parameters is fixed
  - Faster to use but makes stronger assumption about the data distribution
  - Example: Discriminant Analysis (μ, Σ)

- Non-parametric model
  - The number of parameters grows with training data
  - More flexible but computationally intractable for large datasets
  - Example: SVM (the number of support vectors)

Table 8.1

Model	Classif/regr	Gen/Discr	Param/Non
Discriminant analysis	Classif	Gen	Param
Naive Bayes classifier	Classif	Gen	Param
Tree-augmented Naive Bayes classifier	Classif	Gen	Param
Linear regression	Regr	Discrim	Param
Logistic regression	Classif	Discrim	Param
Sparse linear/ logistic regression	Both	Discrim	Param
Mixture of experts	Both	Discrim	Param
Multilayer perceptron (MLP)/ Neural network	Both	Discrim	Param
Conditional random field (CRF)	Classif	Discrim	Param
K nearest neighbor classifier	Classif	Gen	Non
(Infinite) Mixture Discriminant analysis	Classif	Gen	Non
Classification and regression trees (CART)	Both	Discrim	Non
Boosted model	Both	Discrim	Non
Sparse kernelized lin/logreg (SKLR)	Both	Discrim	Non
Relevance vector machine (RVM)	Both	Discrim	Non
Support vector machine (SVM)	Both	Discrim	Non
Gaussian processes (GP)	Both	Discrim	Non
Smoothing splines	Regr	Discrim	Non

Generative/Discriminative pros and cons

- Data fitting
  - Generative: usually very easy, e.g., NB is done by counting and averaging
  - Discriminative: requires solving a convex optimization problem
- Parameter optimization
  - Generative: parameters of class conditional density are estimated independently.
  - Discriminative: interaction among the parameters.
- Feature preprocessing (e.g., replace x with f(x))
  - Generative: hard to define the changes in feature correlation
  - Discriminative: allows input to be arbitrarily preprocessed

Summary

- Generative classifiers make stronger modeling assumptions about the data.
- If the assumptions hold, then generative classifiers fit the data better (better model) with relatively smaller training set.
- For incorrect assumptions, discriminative classifiers perform better since they are less sensitive to incorrect modeling.