Metrics

Machine Learning

Metrics

- Metrics are problem and domain specific
- Find a single number metric to compare models
- Be cautious of metrics others used

Accuracy in classification

- Accuracy measures closeness to true values
- Accuracy = correctly classified / total classified
- Error = misclassified / total classified
- Accuracy is useful in supervised classification problems
- Error is useful in regression problems

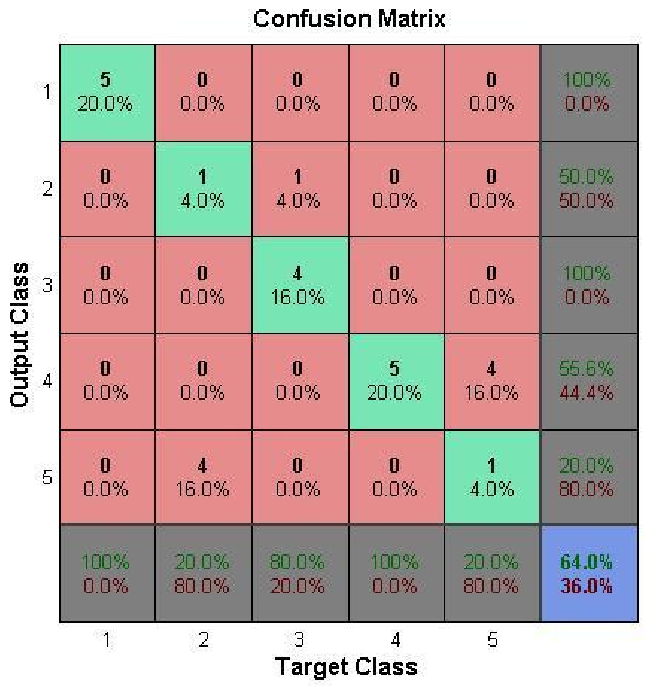
Minimal Accuracy

Higher than chance accuracy

For example 20% accuracy in a classifier with five classes means that classifier is not working.

Confusion matrix

Visualization of where classifier makes mistakes



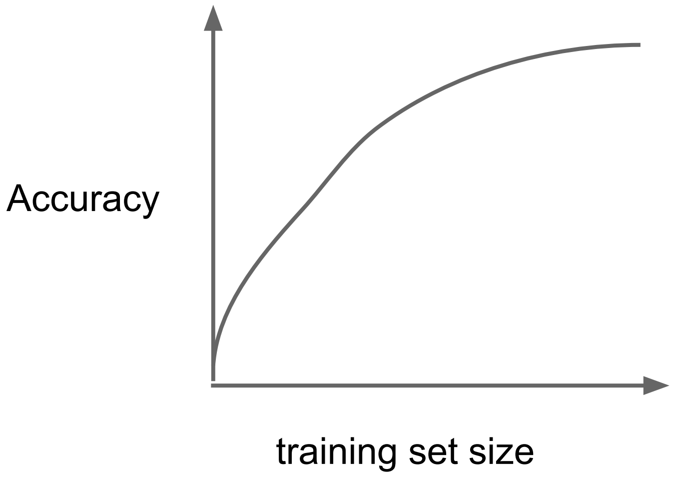
Error

Useful in regression problems

Usually mean squared error:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y_i} - Y_i)^2.$$

Learning curve



- See how much training data you need
- How training improves accuracy
- Sanity check of algorithm
- Spot overfiting

Binary Classifier Metrics

	True label is positive	**True label is negative**
Prediction is positive	True Positive (TP)	False positive (FP)
Prediction is negative	False negative (FN)	True negative (TN)

Accuracy = (TP + TN) / (All)

Recall = TP / (TP + FN)

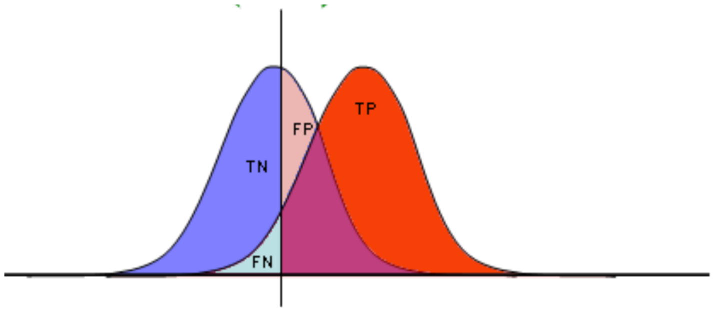
Precision = TP / (TP + FP)

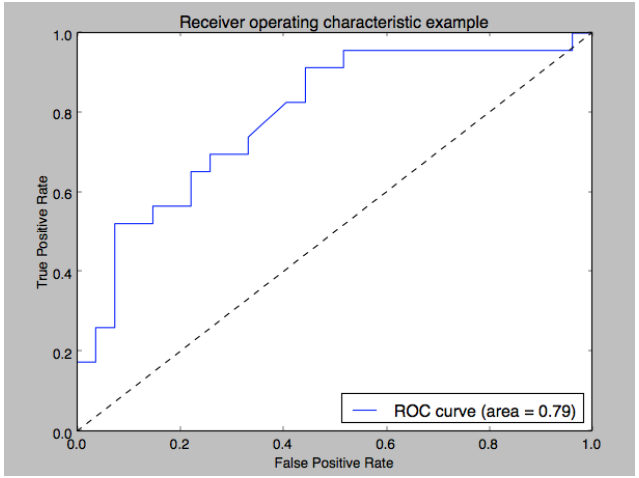
F1 Measure = 2 * (precision * recall) / (precision + recall)

Useful in situations with skewed classes (e.g cancer)

ROC curves

- Receiver operating characteristic
- Visualize performance of binary classifier
- Plots tradeoff between recall and specificity





Validation

Machine Learning

Overfitting

- Big issue in machine learning
- Balance with underfitting
- Happens when test accuracy starts to diverge from training accuracy

Overfitting

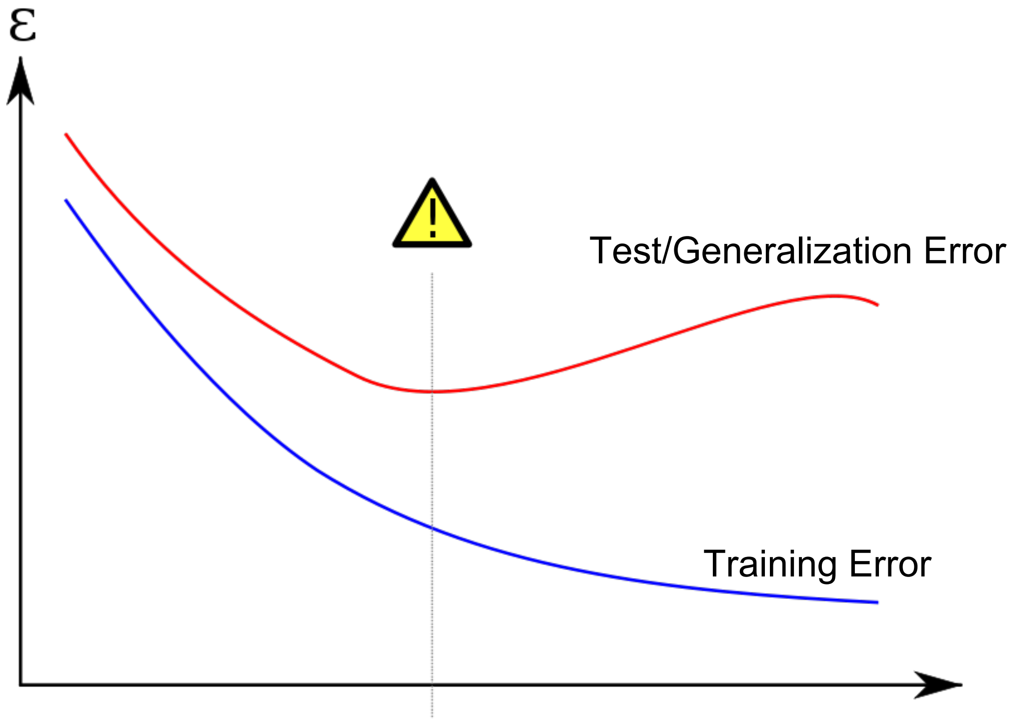


Image from http://en.wikipedia.org/wiki/File:Overfitting_svg.svg

Separating data

Typically split 70% training, 30 % evaluation

Random split

In Sklearn:

```
x_train, x_test, y_train, y_test = cross_validation.train_test_split(data, labels, test_size=0.1, random_state=0)
```

Cross-validation

- Used to detect overfitting
- Cross validation allows use of full data set for training.
- Number of different cross-validation methods
- K-fold is most common

K-fold cross validation

- Divides dataset into K randomly selected subsets
- Each fold one subset is used for validation, and rest for training.
- 10-fold cross validation is common
- Variants: Leave-one-out (LOO), Stratified K-fold

Cross validation in Python

To construct cross validation object

```
kf = cross_validation.KFold(n, n_folds = 10)
```

n = number of elements

n_folds = number of folds

To get the scores

```
cross_validation.cross_val_score(clf, data, labels, kf)
```

Model Selection

Machine Learning

Brute force approach

- Grid search of parameters such as gamma in SVM
- Can take a long time
- Cross-validation is performed on each iteration
- Some score function is computed

AIC / BIC

- Akaike/Bayesian information criterion
- Smaller is better
- Used to find "best" (most appropriate) model
- BIC penalizes complexity more

Likelihood of model

# of data points

$$\mathrm{AIC} = -2\ln L + 2k \qquad \mathrm{BIC} = -2\ln L + k\ln N$$

# of parameters