Feature processing
Create new features

• Constructive induction can be used to try out combinations of features (linear and nonlinear)
  – A very large search space
• Filter out the less useful features. How?
  – t-test, information gain, entropy, gini measure, etc.
Project features

• Random projections
• Principal Components Analysis (PCA)
  – Transform points into a different coordinate system
  – Make sure coordinate system has orthogonal axes.
• Consider examples as a cloud of points
  – Place the first axis in the direction of greatest variance (principal component)
Project features

• The second axis must be perpendicular to the first.
  – Subject to the orthogonality constraint, choose the second access to maximize the variance along it.
  – Continue choosing each axis to maximize its share of the remaining variance such that it is orthogonal to the other axes.
Principal components

• Calculate the covariance matrix of the original coordinates of the points
• Diagonalize this matrix to find the eigenvectors
• These are the axes of the transformed space sorted in order of the eigenvalue (gives the variance along its axis).
• Each axis is a component which accounts for a percentage of the variance.
Principal components

• With the Iris data, the first principal component covers 72.77% of the variance.

• It is:
  0.581petallength + 0.566petalwidth + 0.522sepallength - 0.263sepalwidth

• The second covers 23.03% and is:
  -0.926sepalwidth - 0.372sepallength - 0.065petalwidth - 0.021petallength
Accuracy

• Now 2 features, but only 94% accuracy with J48 and default values in a 10-fold CV.
• Often better accuracy, but 2% worse here with 50% number of features.
• Features clearly matter!