##

Supervised Learning: Support Vector Machine (Linear SVM)
Binary Outcome

Plane can be described by a plane with a normal vector and a scalar b
$w^Tx+b = 0$, w^T (normal vector) is orientation and b (bias) is distance from origin
(normal vector is like the opposite of slope, also has a length)
Width of margin = 2/(||w||) or 2 divided by the length of the normal vector
Goal we have to maximize 2/(||w||) == minimize((||w||)/2)
With the constraints of plus class above -> $w^Tx+b >= 1$
With the constraints of minus class below -> $w^Tx+b <= -1$

This system can be solved by calculus, I will point you to resources if you want to see how

##

Dimension Reduction: t-distributed Stochastic Neighbor Embedding (t-SNE)

Pronounced: tis-knee
Primary purpose is for visualizing high parameter data
It is a form of non-linear dimensionality reduction
Pro: Retains both global and local data structure
Con: Computationally expensive
Con: Data point crowding in 2 dimensional space, sometimes may require clustering algorithm to see fine-grained boundaries between populations

((SEE PPT))

“The goal is to take a set of points in a high-dimensional space and find a faithful representation of those points in a lower-dimensional space, typically the 2D plane. The algorithm is non-linear and adapts to the underlying data, performing different transformations on different regions. Those differences can be a major source of confusion.”

##

Differences Between and why choose which?

Naive Bayes, Gaussian Naive Bayes, Decision Trees, k-Nearest Neighbor, Support Vector Machines, Logistic Regression

http://www.dataschool.io/comparing-supervised-learning-algorithms/

Regression: the output variable takes continuous values. Classification: the output variable takes class labels.
Regression: involves estimating or predicting a response. Classification: is identifying group membership.

Algorithm	**Problem Type**	**Results Human Interpretable**	**Average Predictive Accuracy**	**Training Speed**	**Prediction Speed**	**Performs well with small number of observations?**	**Handles lots of irrelevant features well (separates signal from noise)?**
Naive Bayes	Classification	Somewhat	Lower	Fast (excluding feature extraction_	Fast	Yes	Yes
Decision Trees	Classification/Regression	Somewhat	Lower	Fast (excluding feature extraction_	Fast	Yes	Yes
Random Forest	Classification/Regression	A little	Higher	Slow	Moderate	No	No
kNN	Classification/Regression	Yes	Lower	Fast	Depends on n	No	Yes unless too much noise
SVM	Classification	Somewhat	Lower	Fast	Fast	Yes	No
Logistic Regression	Classification	Somewhat	Lower	Fast	Fast	Yes	No
Neural Networks							