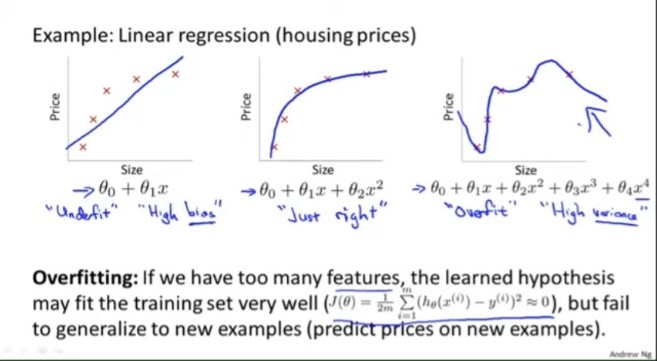
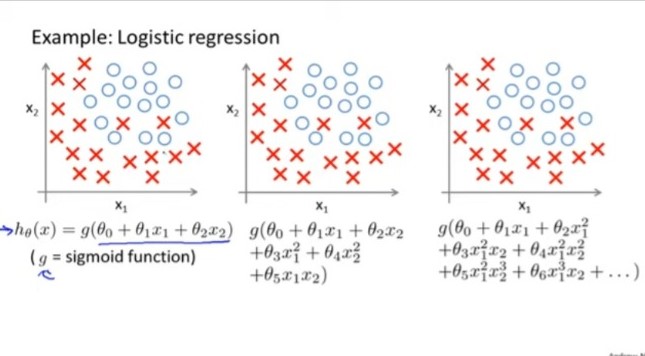
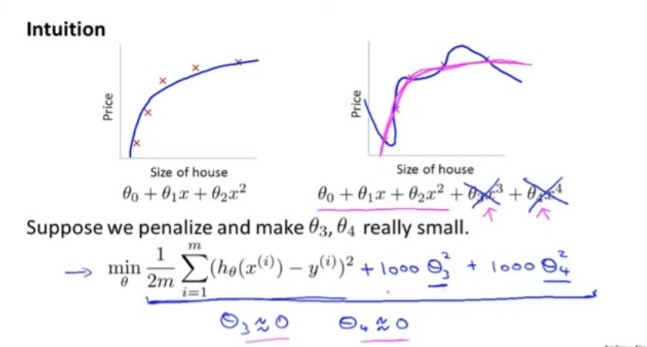
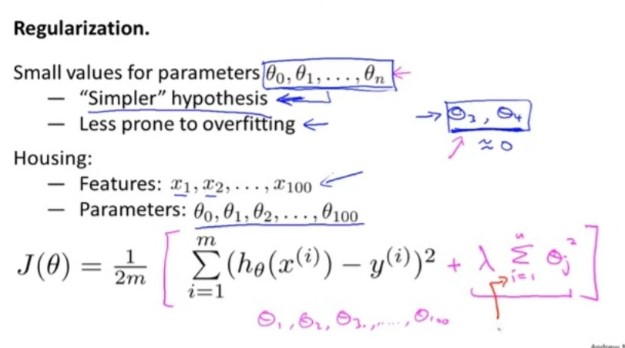
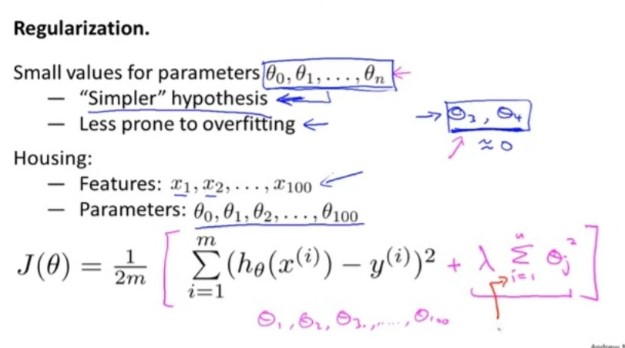
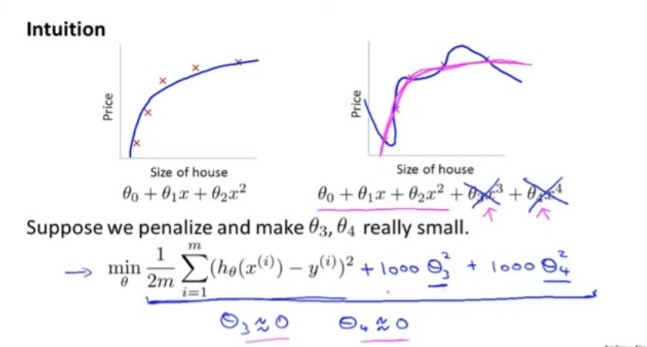
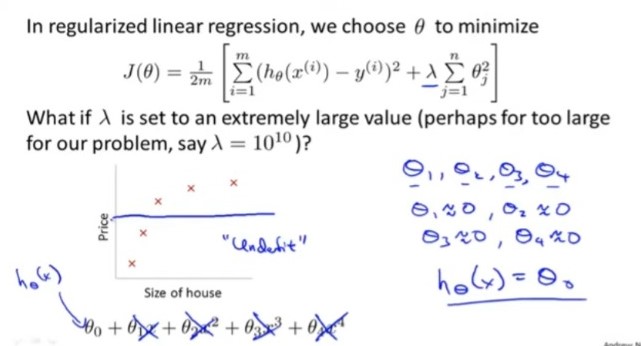
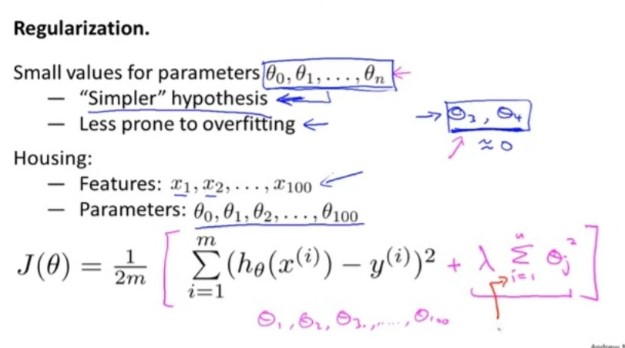
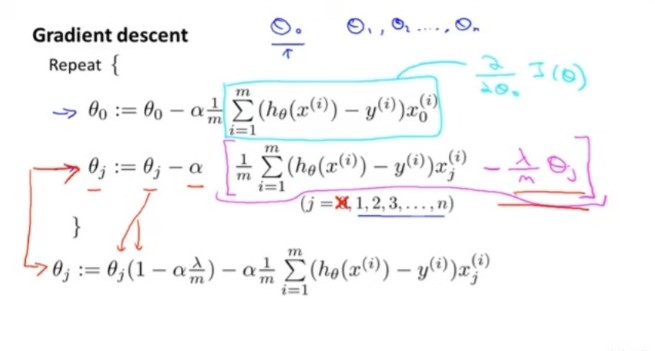
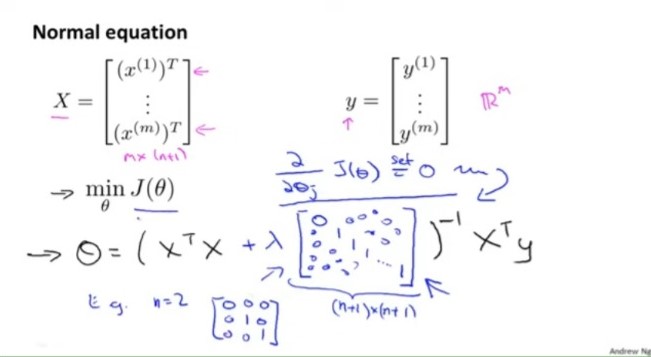
Problem of Over fitting When we have too many features, our hypothesis works well for the training data but fails to predict the output well in the testing data BiasIn machine learning, bias refers to the difference between the predicted outputs of a model and the actual outputs, averaged over many different data samples. A model with high bias is one that is too simple and makes strong assumptions about the data. A model with low bias is one that is complex and flexible enough to capture the underlying patterns in the data.High bias models tend to underfit the data, Low bias models tend to overfit the data, meaning they capture the noise in the data and have high variance.High Variance – Works well in Training data But the difference between Predicted output and training output is high Low Variance – Works well in Training data But the difference between Predicted output and training output is low (good prediction result)To Overcome Overfitting Reduce number of Features or Manually select featuresModel Selection Algorithm \* Keep all the features but reduce the parameter valuesWorks well when we have too many features ,each contributes a bit to predicting y

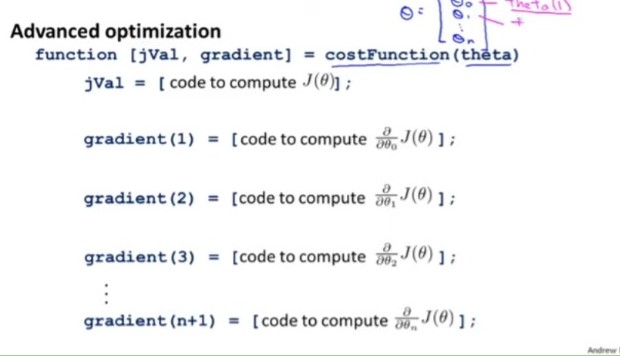
REGULARIZATIONRegularization is a technique used in machine learning to prevent overfitting, which is when a model fits the training data too closely and does not generalize well to new data. Regularization adds a penalty term to the objective function of the model, which helps to control the complexity of the model and reduce the effects of noise or fluctuations in the training data.There are several different types of regularization techniques, including:L1 regularization: Also known as Lasso regularization, this technique adds a penalty term proportional to the absolute value of the model parameters. L1 regularization can be used to perform feature selection, as it tends to set some of the model parameters to zero.L2 regularization: Also known as Ridge regularization, this technique adds a penalty term proportional to the square of the model parameters. L2 regularization tends to distribute the weight of the model parameters more evenly and can be more stable than L1 regularization.Elastic Net regularization: This technique combines L1 and L2 regularization, adding a penalty term proportional to both the absolute value and the square of the model parameters.Regularization can be applied to a wide range of machine learning models, including linear regression, logistic regression, and neural networks. The strength of the regularization can be controlled using a hyperparameter, which can be tuned using cross-validation to find the best value for the given dataset.In summary, regularization is a powerful technique for controlling the complexity of machine learning models and preventing overfitting. It can be applied to a wide range of models and is controlled using a hyperparameter that can be tuned using cross-validation.Top of Formƛ is called Regularization Parameter which maintains the trade off between fitting the hypothesis and maintain penalize term.

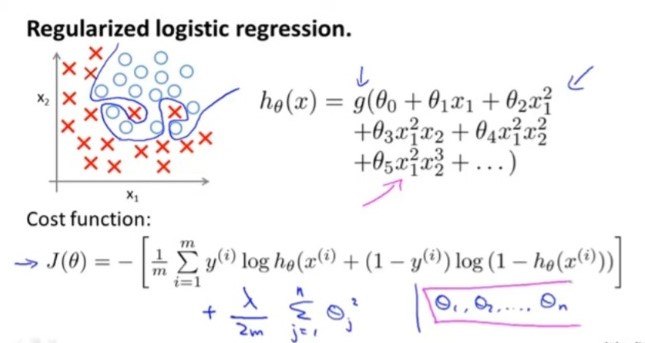
Regularized Linear Regression





Regularized Logistic Regression



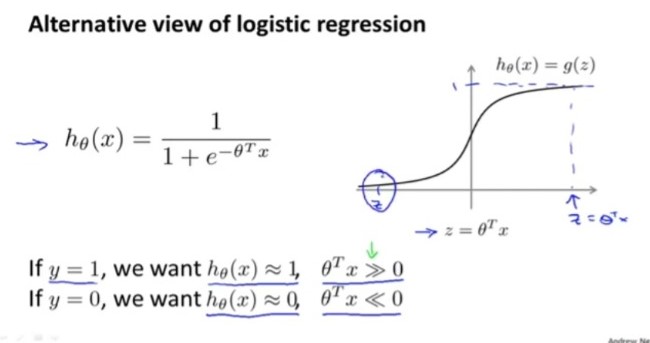


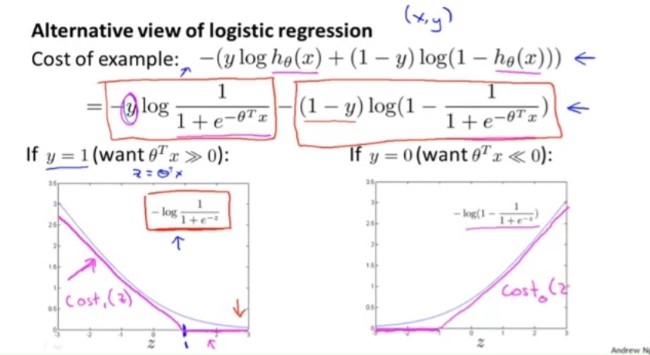
SUPPORT VECTOR MACHINE

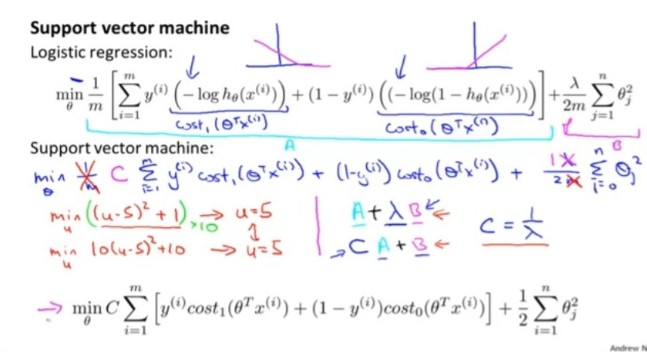
Optimized version of Logistic regression

Largin Margin Classifiier for Linear data

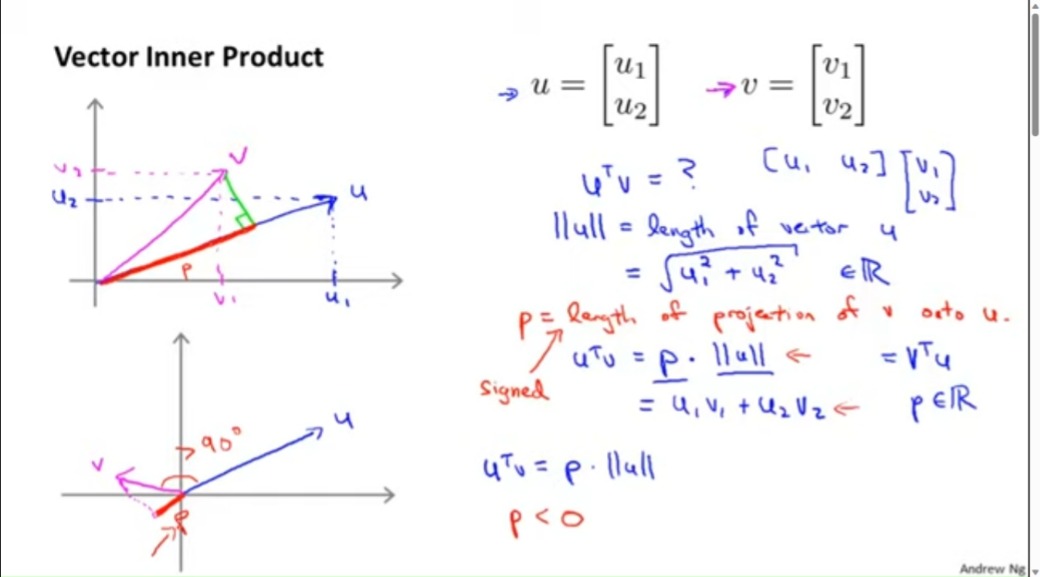
Kernel method for Non Linear data

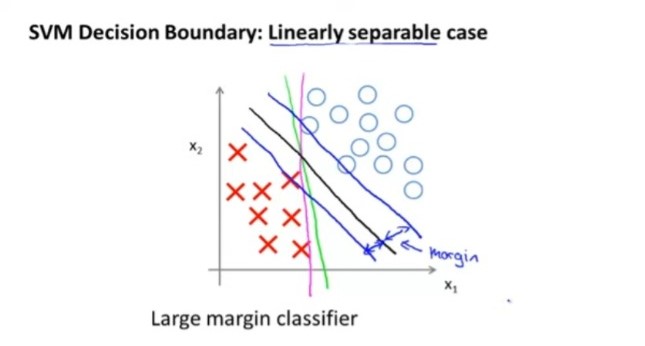
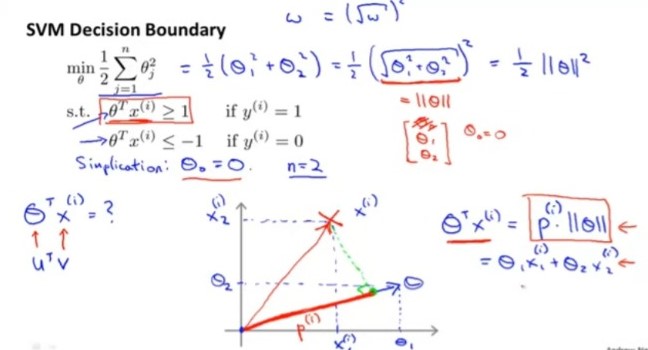


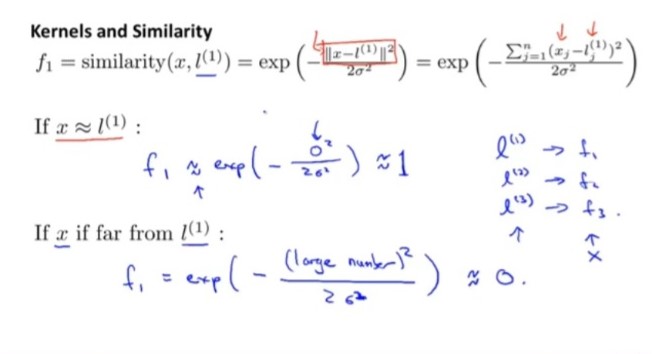


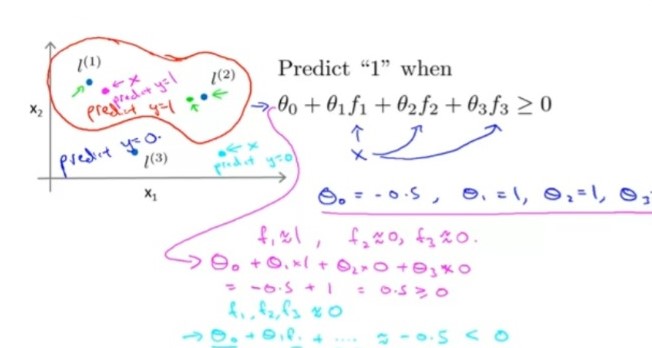


SVM LARGE MARGIN INTUITION

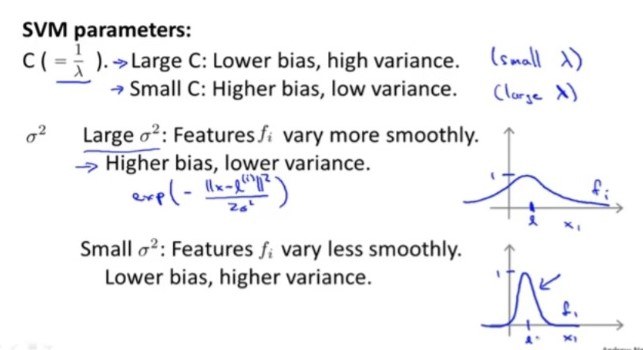
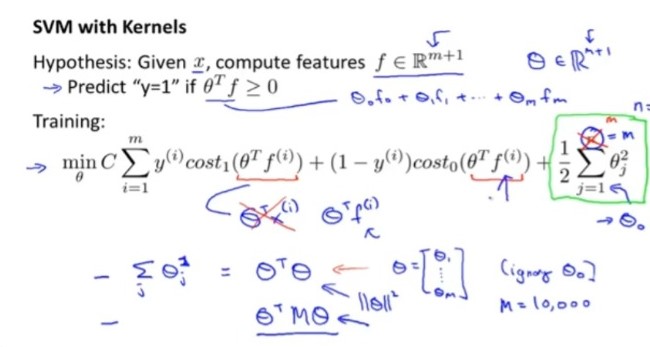
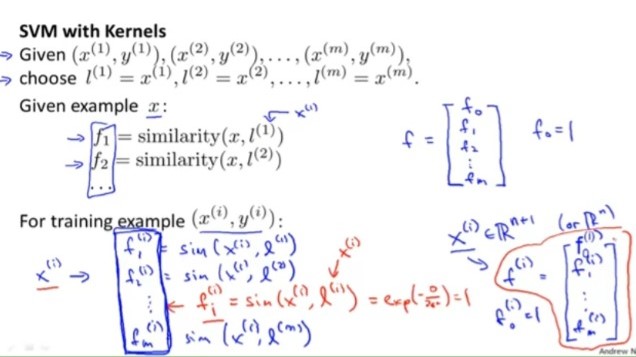


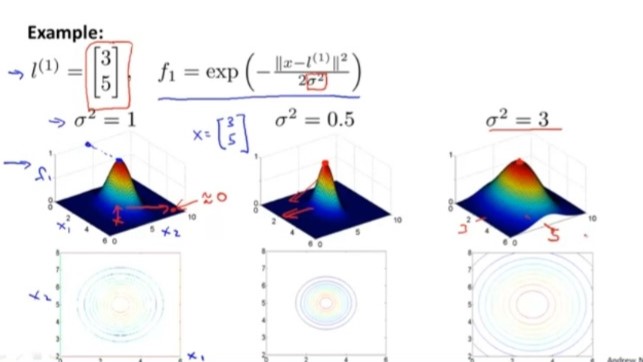


KERNEL



KERNEL 2





Clustering

Unsupervised learning technique which groups data based on inherent characteristics or Similarities.The goal of clustering is to identify natural groupings or clusters within a dataset, where data points within the same cluster are more similar to each other than to those in other clusters. It can be used for various purposes, such as pattern recognition, data exploration, outlier detection, and data compression.

Algorithms

1 K means Clustering

2 Hierarchal Clustering

3 Gaussian Mixture Model etc

K means Clustering

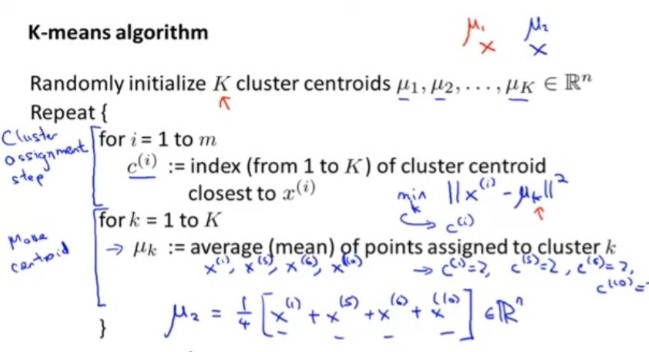
It is a Iterative algorithm which partition dataset int K clusters where K is a user defined parameter Each cluster is represented as centroid the algorithm iterative assign data to the nearest cluster and update the cluster by mean/Average of Dat points of particular cluster unti convergence(Cluster positon not change for aparticular mean)

Inputs

K cluster

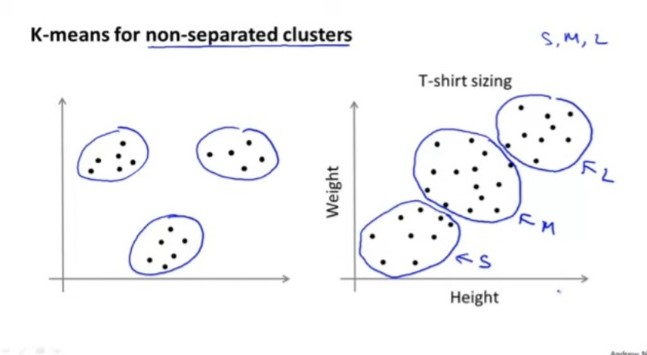
X training set{x1,x2…xn}

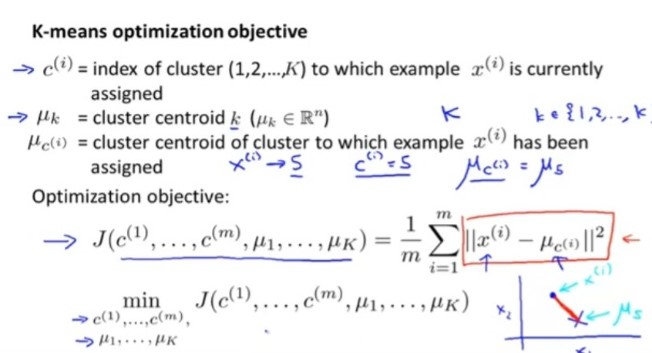
X(i) belongs t o R(n) (drop x0 =1 for convention)



Note : Min k uses squared distance instead of normal distance by convention where both give same minimum

T SHIRT SIZING EXAMPLE





**Random Initialization**

\*should k<m

\*Pick K random data points from dataset

\* Assign k clusters as that data points

Best Practice

Iterate from 1 to 100

Randomly initialize K cluster centroids

Run K means

Compute cost function(distortion)

Pick clusters with minimum Distortion

