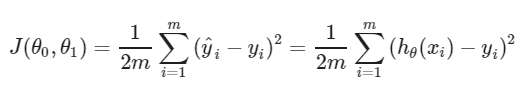
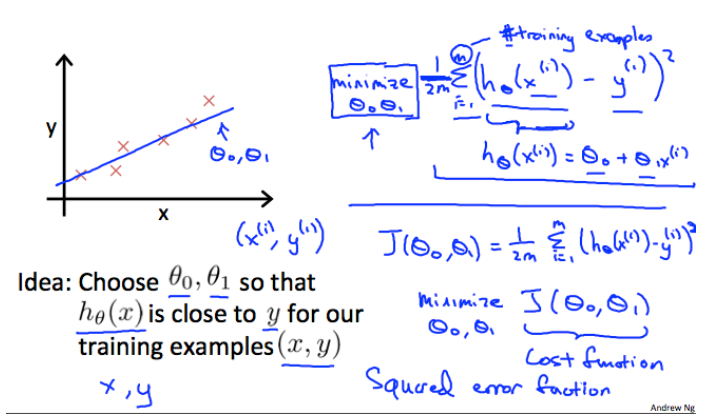
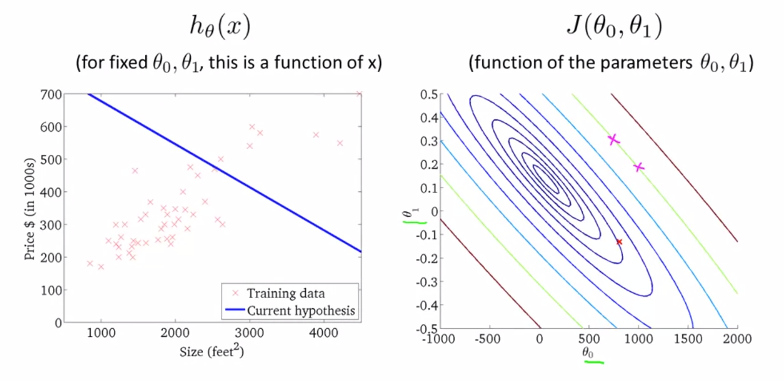
**Week 1:**

1. Algo and math not just enough, application is imp
2. ML grew out of AI
3. Examples:
   1. Click stream data used know more about users
   2. Medical record to medical knowledge
   3. Genomics
   4. Autonomous helicopter (can’t be programmed)
   5. NLP
   6. Handwriting recognition
   7. CV
   8. Recommendation systems
4. ML – ability to learn without being explicitly programmed (Checkers example)
5. Supervised Learning – labelled dataset (for every data point we had “right” answer)
   1. Classification / Regression
6. Unsupervised Learning – don’t give “right” answers
   1. Clustering
      1. Google news (cohesive news clustering)
      2. Clustering people into group from gene data
      3. Social network analysis
      4. Market segmentation
   2. Cocktail Party problem – hard to hear people in noisy party
      1. Multiple speaker recording multiple people’s audio
      2. Pass them to algo, then algo will split the audios
7. Model Representation – Regression Supervised problem (hypothesis)
8. Cost Function – Help fit best line to data (choosing parameters of model changes hypothesis function)
   1. Linear regression – we need to determine 2 parameters θ0 and θ1 so as to minimize squared error of all the training examples

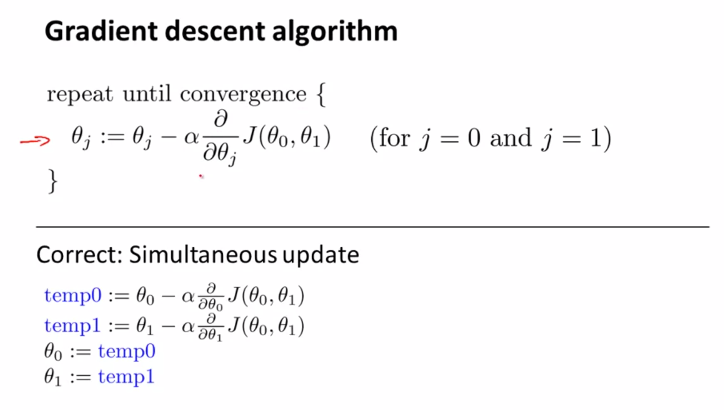




* 1. Contour plots used for better visualizing cost function minima point

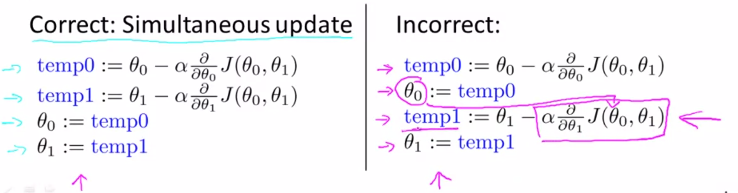


1. Gradient Descent – used to minimize cost function (in linear regression). It applied to any general cost function
   1. Vary θ0 and θ1 and go towards local minima



Alpha – learning rate, derivative of cost function (slope of tangent)

* 1. Correct way to implement gradient descent (simultaneous update)

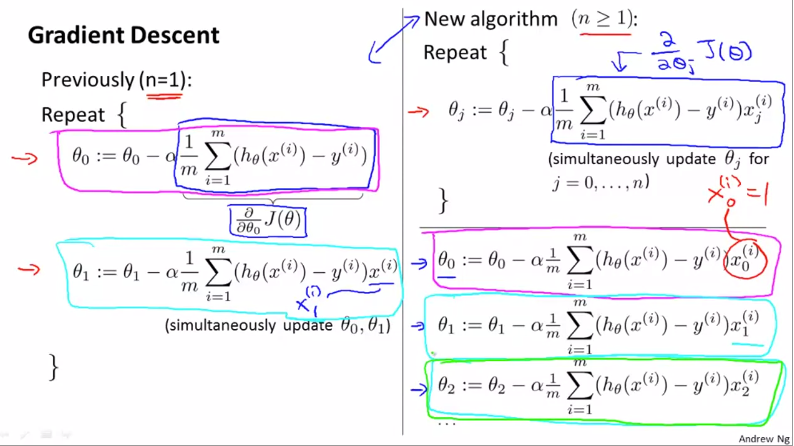


* 1. For linear regression, we will always have a convex function for cost

1. Linear Algebra:
   1. Matrix: dimensions
   2. Vector: n x 1 matrix, 1 indexed (1,2, 3….), 0 indexed (0,1, 2…)
   3. Matrix Algebra: addition, scalar multiplication, vector multiplication
      1. Express linear regression hypothesis function as vector multiplication
      2. Prediction = data matrix \* parameters
      3. Method to apply multiple hypothesis on dataset (linear regression)
   4. Matrix properties:
      1. A X B not equal to B X A
      2. A X (B X C) = (A X B) X C
      3. A X I = I X A = A
      4. Matrix which don’t have inverse are singular/ degenerate

**Week 2:**

1. Multiple Features (Multivariate linear regression):
   1. n = number of features, h depends on θ0 θ1 θ2 θ3
   2. To accommodate the constant term in the equation of hypothesis, theta matrix is represented in 0 indexed notation consisting of n+1 term
2. Gradient Descent for multiple variables:



1. Feature Scaling:
   1. Convergence is fast if all features are in similar scale otherwise, we might get skewed contours
   2. It’s better to scale them in 0 to 1 (or -1 to +1) which will help in reaching global minimum faster
   3. Mean normalization: subtract all values with mean (and divide by range)