Minimzing MSE for Linear Models

BACK TO. 2-5 NOTES PAGE 2 Complexity of solving equations or finding matrix inverse is $O(Nd^2 + d^3)$

Minimizing MSE for Non-Linear Equations

$MSE(\theta)$ is concave so you can use gradient descent:

$$\theta^{new} = \theta^{current} - stepSize * gradient(MSE)$$

Stochastic Gradient Descent can also be used, faster but noisier

Probabilistic Interpretation of Regression

$p(y|x)$: for fixed x, there is variation in y 2 types of variation: - Measurement noise - Unobserved Variables 2 sources of variability: $p(y|x)$: variability in y given x $p(x)$: distribution of input data in the space We have a joint distribution: $p(x, y) = p(y|x)p(x)$ and we learn $p(y|x)$

Modeling Framework

$y_x = E[y|x] + e$ where y_x: what we observe $E[y|x]$: what we try to learn with f(x,theta) e: unpredictable error term

Simple Model:

$p(y|x) = N(f(x, theta), sigma^2)$ where $f(x, theta) = theta^T x$

Conditional Likelihood for Regression

$$L(\theta) = \prod p(y_i|x_i, \theta)$$