Introduction to Machine Learning

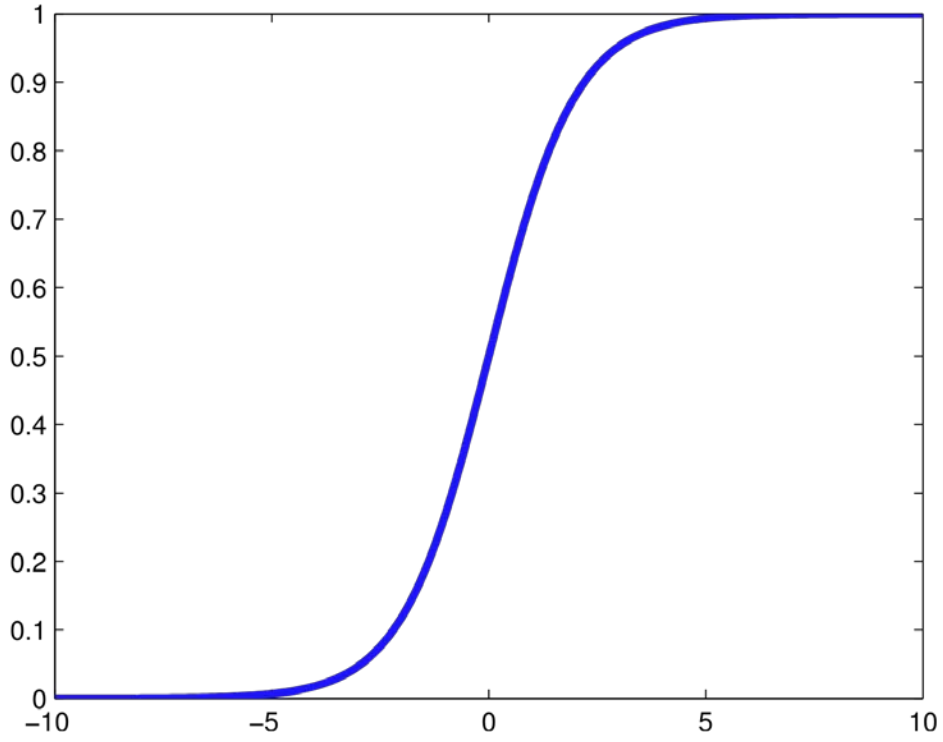
Brown University CSCI 1950-F, Spring 2011
Prof. Erik Sudderth

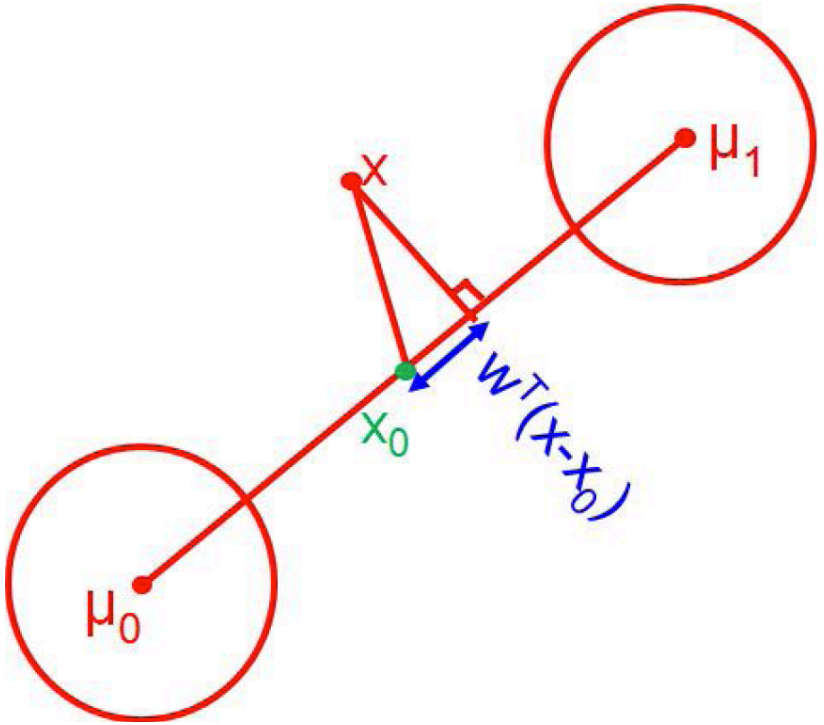
Lecture 11: Stochastic Gradient Methods, Iteratively Reweighted Least Squares

Many figures courtesy Kevin Murphy's textbook,
Machine Learning: A Probabilistic Perspective

Binary Discriminant Analysis

Logistic Function





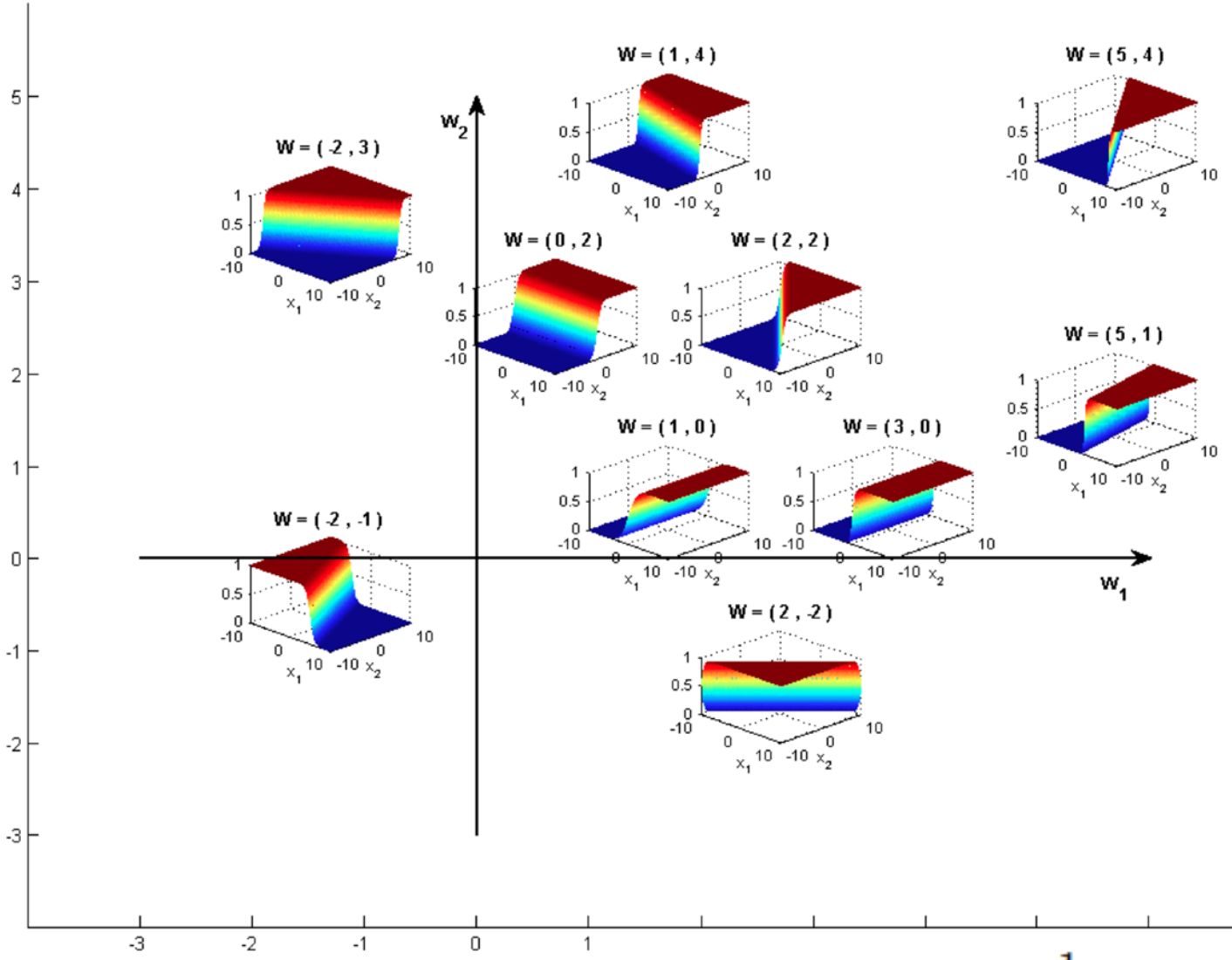
$$\text{sigm}(\eta) := \frac{1}{1+\exp(-\eta)} = \frac{e^{\eta}}{e^{\eta}+1}$$

$$p(y=1|\mathbf{x},\boldsymbol{\theta}) = \sigma(\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0))$$

$$\mathbf{w} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0 = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\frac{\log(\pi_1/\pi_0)}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}$$

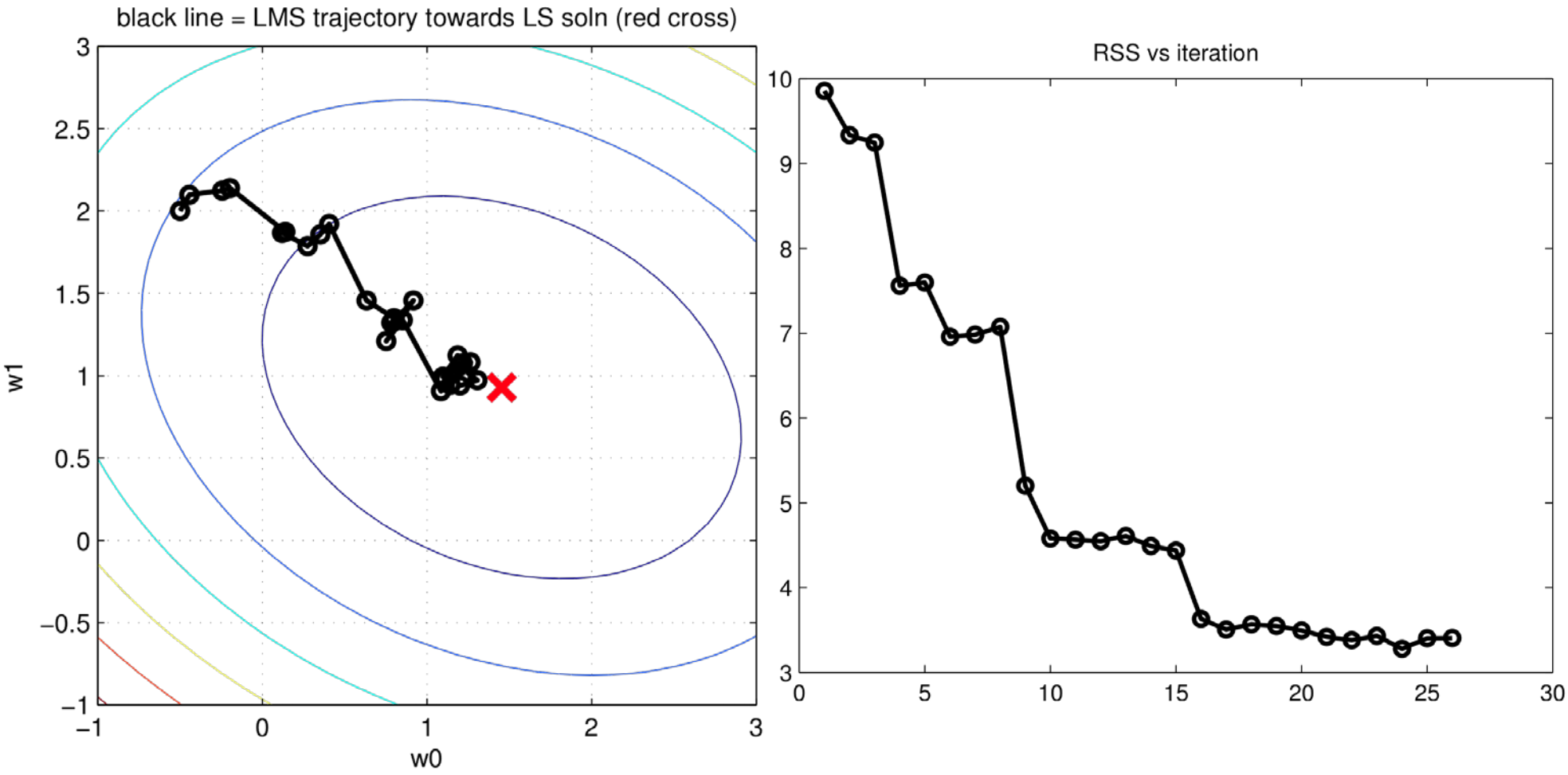
Logistic Regression



$$p(y|\mathbf{x}, \mathbf{w}) = \mathrm{Ber}(y|\mathrm{sigm}(\mathbf{w}^T\mathbf{x}))$$

$$\mathrm{sigm}(\eta) := \frac{1}{1+\exp(-\eta)} = \frac{e^{\eta}}{e^{\eta}+1}$$

Least Mean Squares (LMS)



Stochastic gradient descent applied to linear regression model