

An aerial photograph of a city, likely Madison, Wisconsin, taken from a high vantage point looking down at the city and the surrounding water. The sun is setting behind a hill in the background, creating a warm, golden glow over the entire scene. The city is densely packed with buildings, and the water is filled with numerous sailboats. A large, semi-transparent rectangular box is overlaid on the center of the image, containing the course title in large, bold, black text.

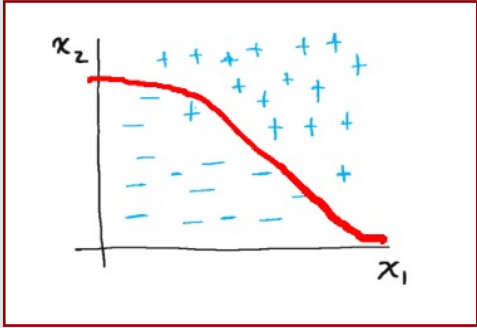
CS/ECE/ME 532 Matrix Methods in Machine Learning

Welcome!



Activity 8

Binary Linear Classifier



$$\begin{aligned} \text{sign}(\mathbf{x}_1^T \mathbf{w}) &= -1 \\ \text{sign}(\mathbf{x}_2^T \mathbf{w}) &= +1 \end{aligned}$$

Lots of training data

Drop "sign"

$$Xw \approx y$$

Any downsides to the squared error loss function?

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2$$

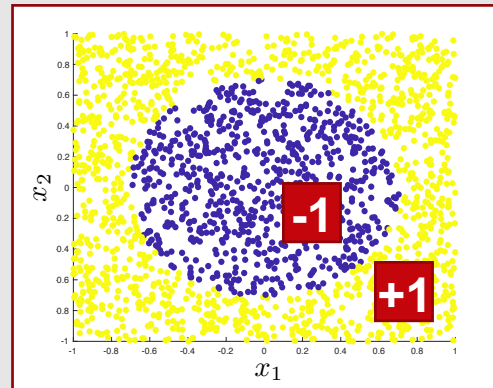
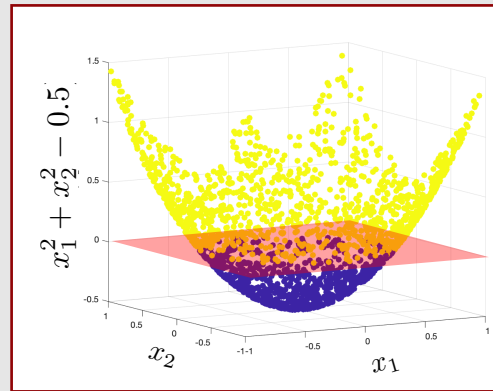
$$w^* = (X^T X)^{-1} X^T y$$

Curved boundaries

$$\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & 1 & (x_1^2 + x_2^2) \end{bmatrix}$$

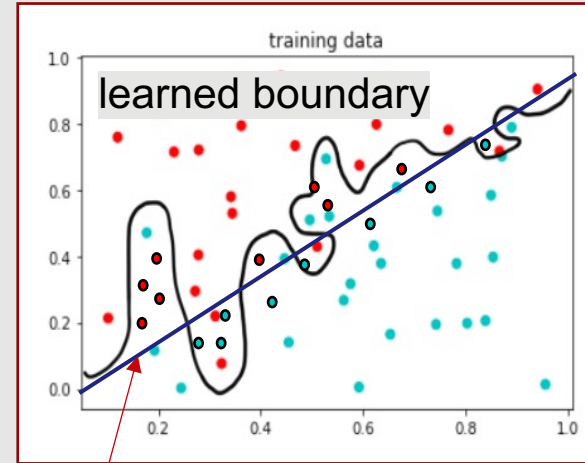
$$\mathbf{w} = (0, 0, -0.5, 1)$$

$$\Rightarrow \hat{y} = \text{sign}(x_1^2 + x_2^2 - 0.5)$$

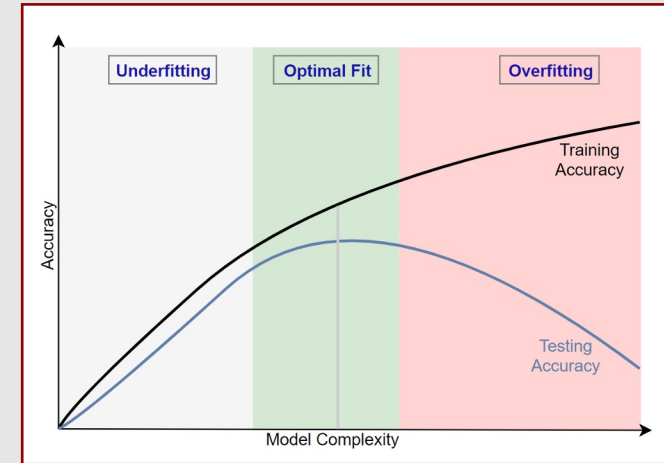


(Bad) idea: Keep adding many powerful non-linear features!

- Leads to **overfitting**
 - High training performance, poor testing performance



true boundary



Underfitting is also undesirable!

How can we find the “right” number of non-linear features?

Cross-validation:

