

An aerial photograph of a city, likely Madison, Wisconsin, taken from a high vantage point. The city is situated on a peninsula, with a large body of water (Monona Lake) in the foreground and background. The sun is setting behind a hill in the distance, creating a warm, golden glow over the entire scene. Several sailboats are visible on the water. A large, semi-transparent rectangular box is overlaid on the center of the image, containing the course title in black text.

CS/ECE/ME 532 Matrix Methods in Machine Learning

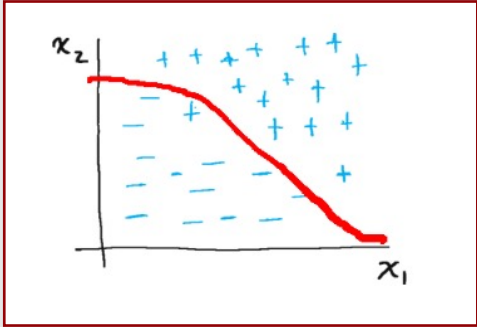
Welcome!



Activity 8



Binary Linear Classifier



$$\text{sign}(x_1^T w) = -1$$

$$\text{sign}(x_2^T w) = +1$$

$$\vdots$$

Lots of training data

Drop "sign"

$$Xw \approx y$$

(no solution)

Any downsides to the squared error loss function?

$$w^* = \arg \min_w ||Xw - y||_2^2$$

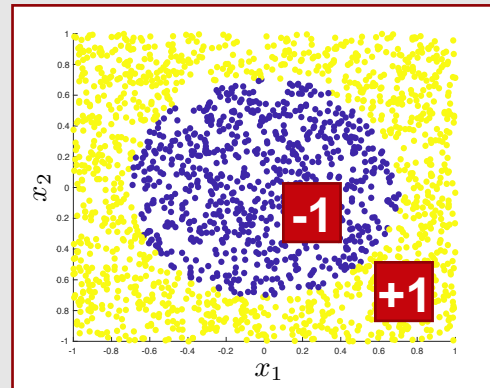
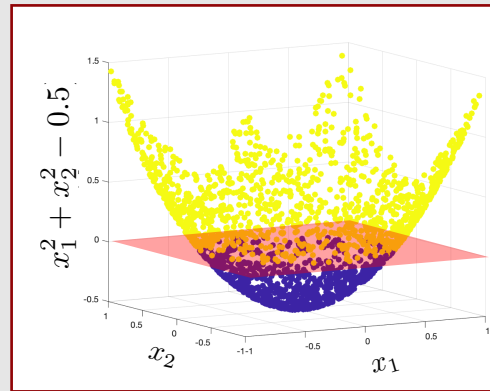
$$w^* = (X^T X)^{-1} X^T y$$

Curved boundaries

$$x^T = [x_1 \quad x_2 \quad 1 \quad x_1^2 + x_2^2]$$

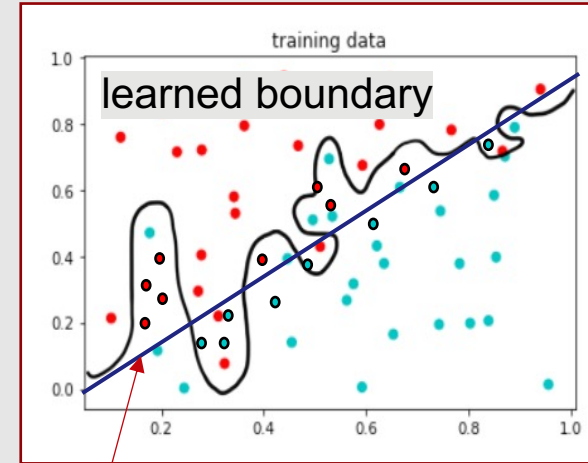
$$w = (0, 0, -0.5, 1)$$

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

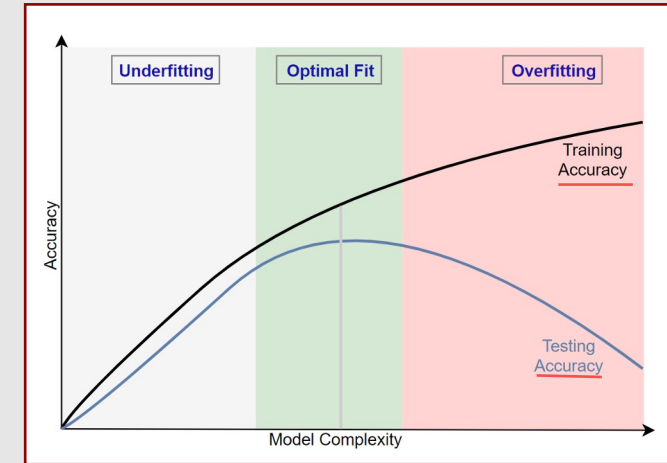


(Bad) idea: Keep adding many powerful non-linear features!

- 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:

