Classifying Data and Matrix Multiplication

Objectives

- introduce the classification problem
- define linear classifiers
- write supervised learning of linear classifiers using matrices

Classification is assigning a category to data

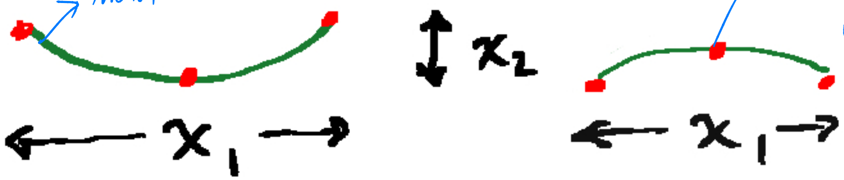
Example: Decide  vs 

"happy" "sad"

- extract features based on a "model"

mouth wider + corners above the center

mouth narrower + corners below the center

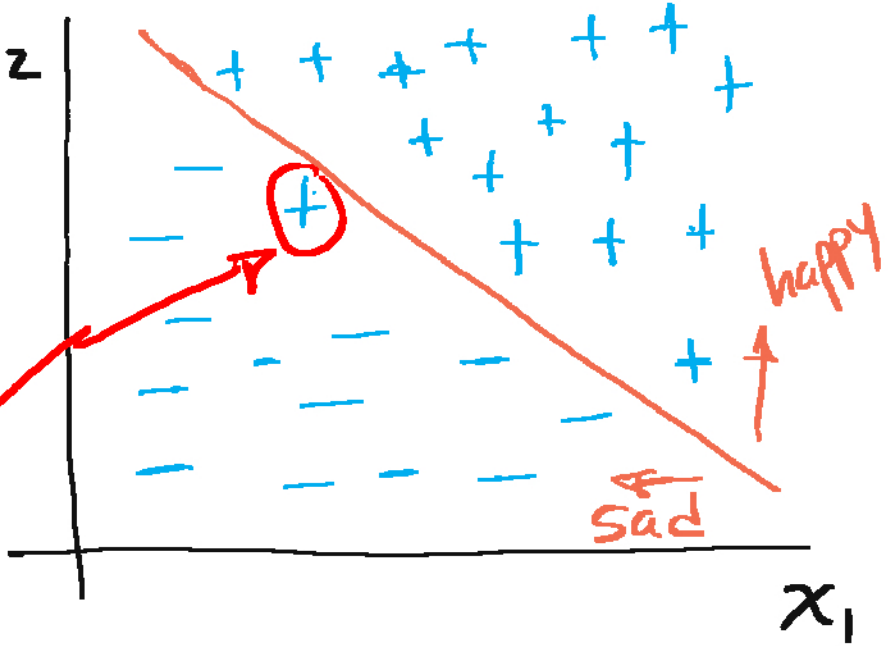


- find a decision boundary

+: happy

−: sad

misclassification



If we use a line to separate classes

$$x_2 = m x_1 + b \Rightarrow x_2 - m x_1 - b = 0$$

Find m, b

Rewrite boundary as an inner product

$$\underbrace{\begin{bmatrix} x_2 & x_1 & 1 \end{bmatrix}}_{\text{feature}} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \underline{x}^T \underline{w} = 0$$

classifier weights

Related to the slope and intercept

☆ Find $\underline{w}$.

Curved decision boundaries:

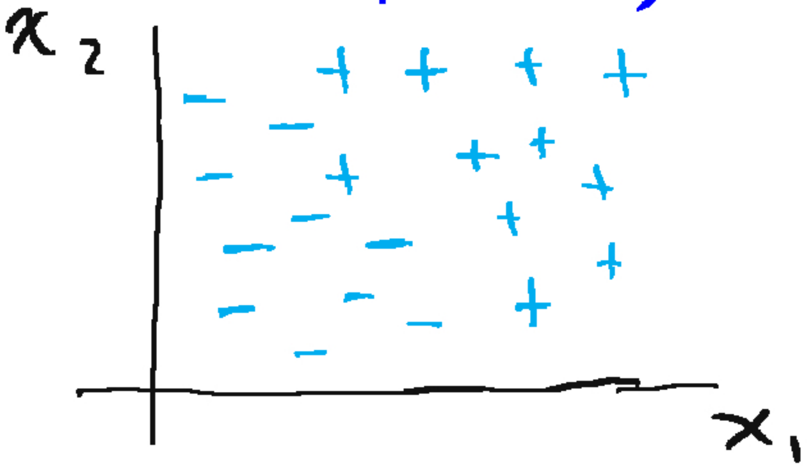
$$\underbrace{\begin{bmatrix} x_2 & x_1^3 & x_1^2 & x_1 & 1 \end{bmatrix}}_{\text{feature}} \begin{bmatrix} w_1 \\ \vdots \\ w_4 \\ w_5 \end{bmatrix} = \underline{x}^T \underline{w} = 0$$

Find $\underline{w}$

A linear classifier is based on a weighted sum of features $\underline{x}^T \underline{w}$

features → $\underline{x}$; weights → $\underline{w}$

Labels specify class associated with a feature



+ happy : label "1"

− sad : label "−1"

binary classification

Supervised learning: given features/labels

$\star\ (\underline{x}_i, l_i)$ find $\underline{w}$ so $\underline{x}_i^T \underline{w} \approx l_i$

Training a linear classifier involves solving a system of linear equations

$$\begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix} \underline{w} \approx \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_N \end{bmatrix} \Rightarrow \underset{N \times M}{\underline{X}}\ \underset{M \times 1}{\underline{w}} \approx \underset{N \times 1}{\underline{l}}$$

N sets of data

N training samples

M features → M个重要特征

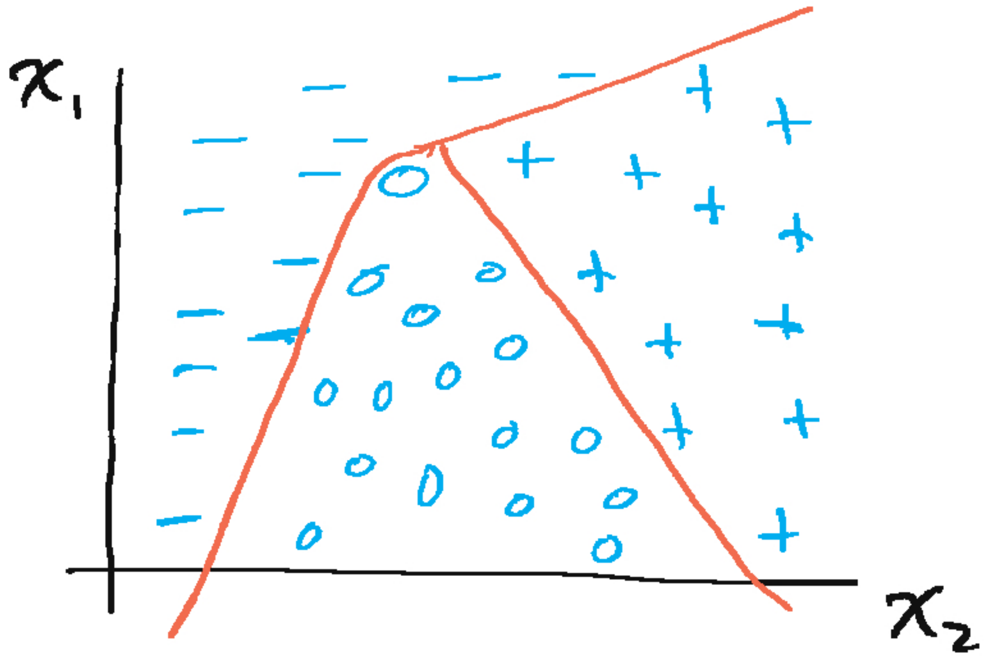
we already had the data → predict new data

Classify candidate feature $\tilde{\underline{x}}$ using $\underline{w}$

if $\tilde{\underline{x}}^T \underline{w} > 0 \Rightarrow$ label "1", if $\tilde{\underline{x}}^T \underline{w} < 0 \Rightarrow$ label "-1"

Advanced topics

- What should we choose for "≈"? $\underline{X}\,\underline{w} \approx \underline{l}$ ★ (How do we measure the closeness between Xw and known labels)
- Feature choice / boundary complexity
- Performance evaluation
- M-ary classification



canvas.wisc.edu/courses/465084/quizzes/663155

Wrong answer

Question 4

0 / 1 pts

Let $\boldsymbol{x}$ be a be a vector that is a function of the features. A binary classifier written in the form $\text{sign}(\boldsymbol{x}^T \boldsymbol{w})$ cannot have a curved decision boundary in the feature space.

why? → Although x is a function of features, they may not be original features

e.g. ($x = \begin{bmatrix} x^2 \\ x \\ 1 \end{bmatrix}$)

★ This forms a parabolic boundary in feature space

True

False

Correct answer