

LECTURE 7

SPRING 2021

APPLIED MACHINE LEARNING

CIHANG XIE

SLIDE CREDIT:

NARGES NOROUZI

COMPSCI 188@UC BERKELEY

HOMEWORK

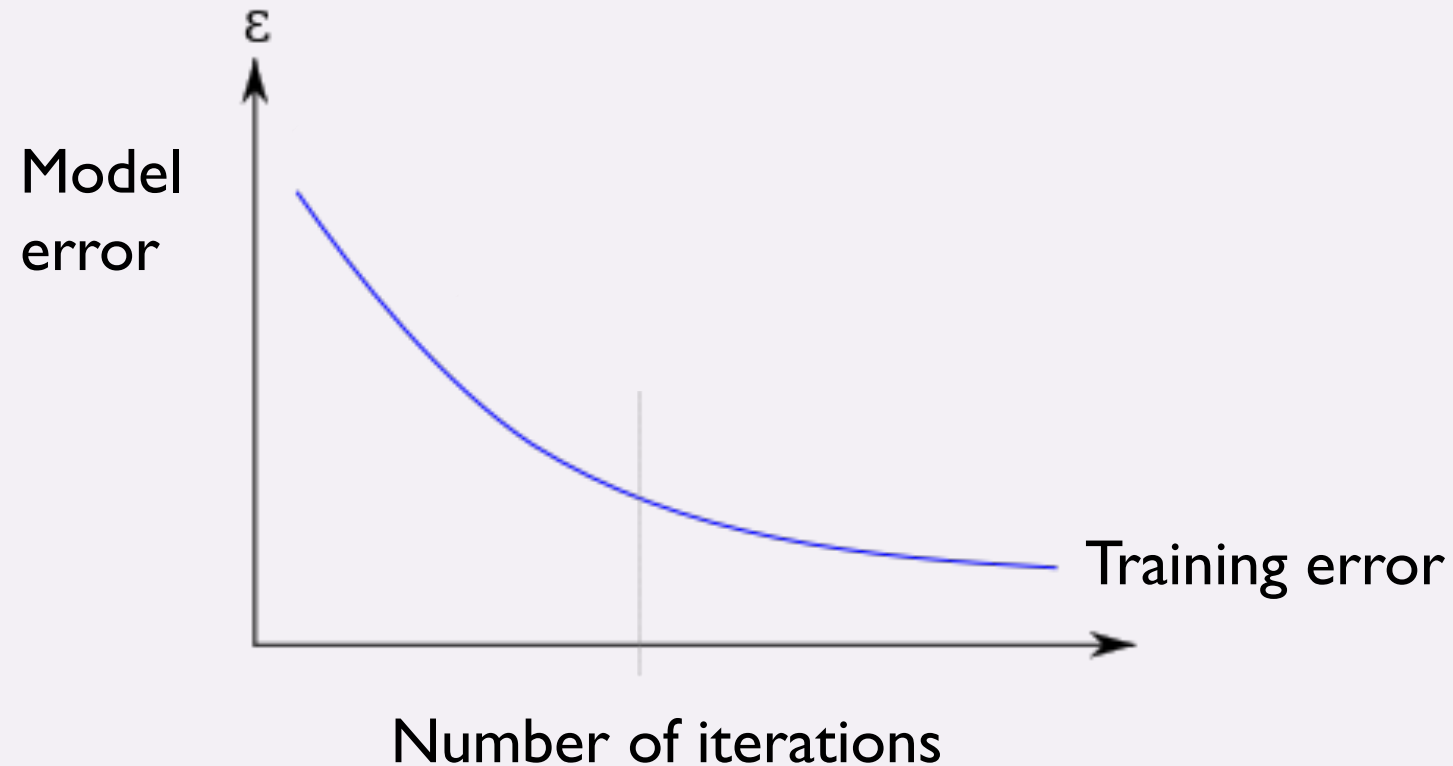
- The efficiency of code will not be considered
- The performance on test set will not be considered
- [minor] For Q1, please do the “real” split on the training set; rather than just provide the index of splitted dataset

TODAY

- Review of Overfitting & Regularization
- Linear Classifiers
 - Binary vs. multi-class
 - Deterministic
 - Probabilistic

OVER-FITTING

Over-fitting during training



REGULARIZATION

- Linear regression objective function

$$Cost(\theta) = \left\{ \frac{1}{2 \times n} \sum_{i=1}^n (h_{\theta}(x^{(i)}) - y^{(i)})^2 \right\}$$

Model fit to data

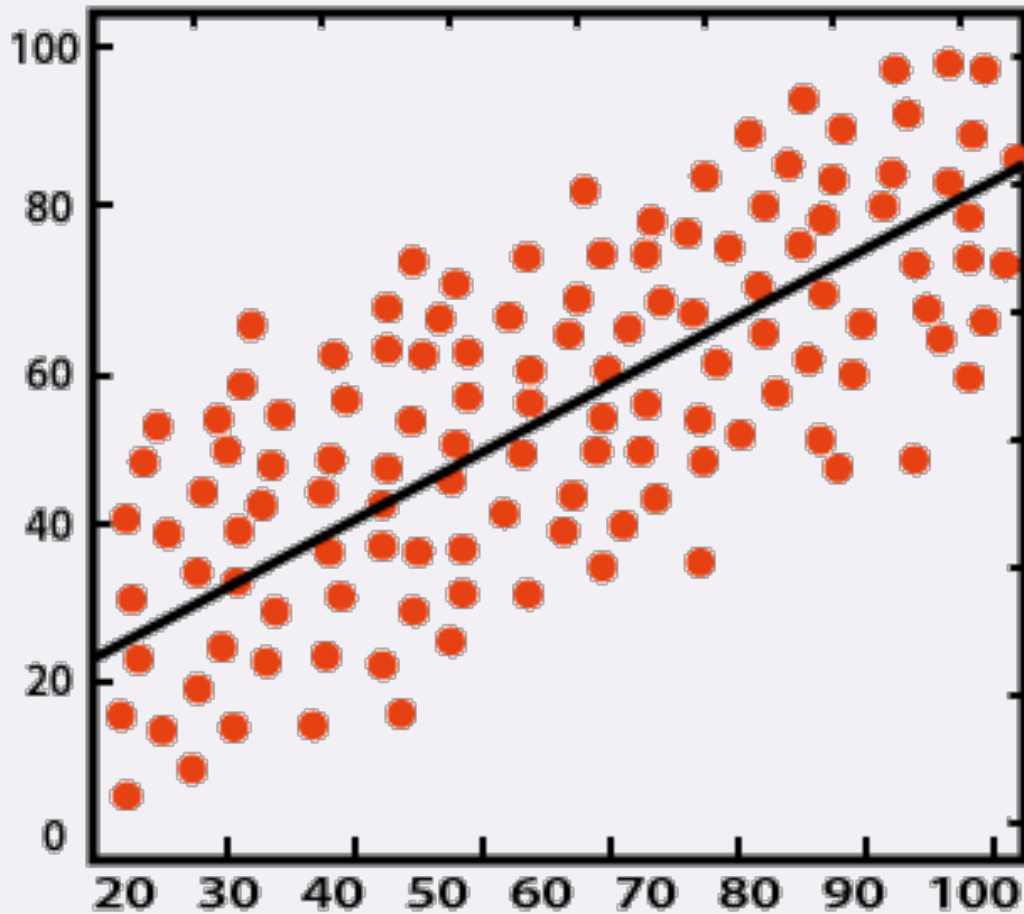
Regularization

- λ is the regularization parameter ($\lambda \geq 0$)
- No regularization on θ_0

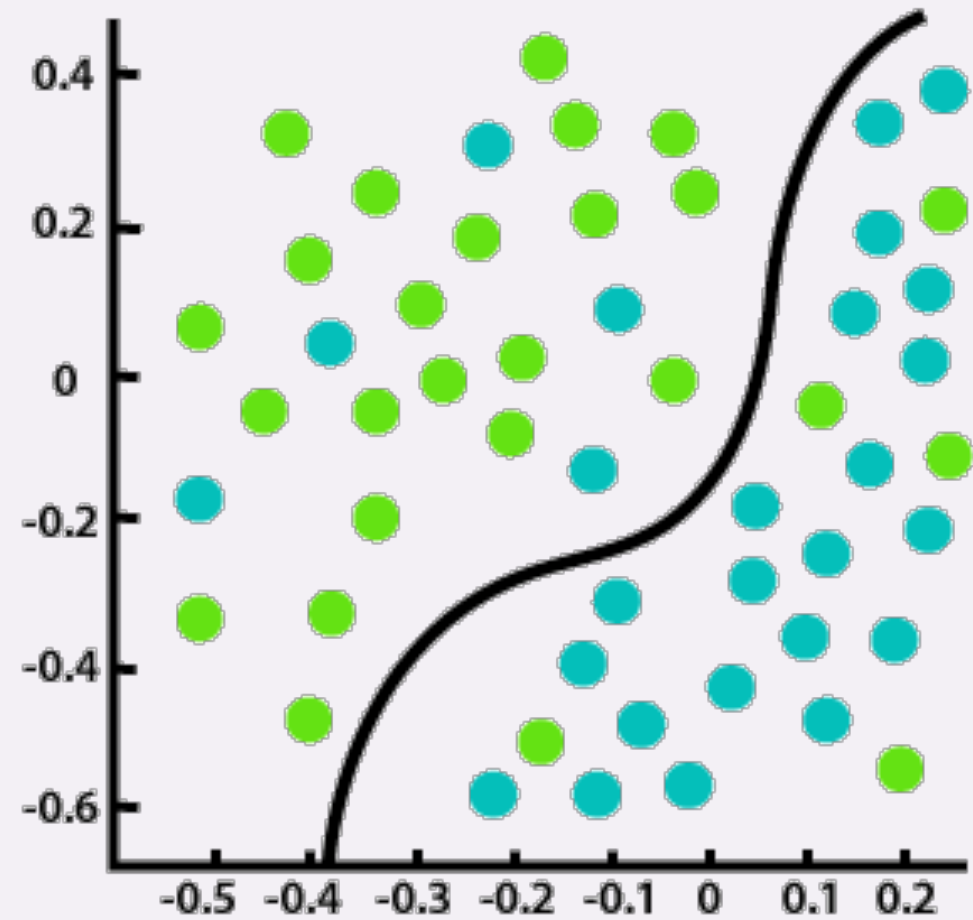
A decorative purple wavy line runs vertically along the left side of the image, starting from the top and extending to the bottom. It has a thick, irregular, hand-drawn appearance.

L2 REGULARIZATION PLAYGROUND

[HTTPS://BIT.LY/32RSU2T](https://bit.ly/32RSU2T)

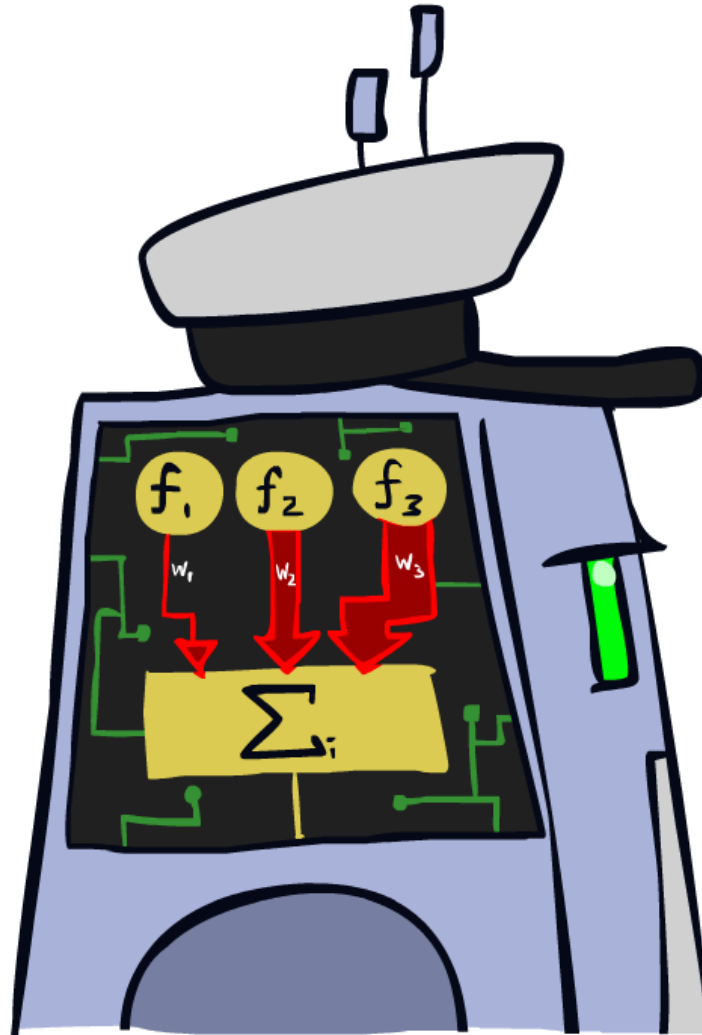


Regression



Classification

LINEAR CLASSIFIERS



FEATURE VECTORS

 x

```
Hello,  
  
Do you want free printer  
cartridges? Why pay more  
when you can get them  
ABSOLUTELY FREE! Just
```

 $\phi(x)$

```
# free      : 2  
YOUR_NAME   : 0  
MISPELLED   : 2  
FROM_FRIEND : 0  
...
```

 y

SPAM
or
NOT SPAM



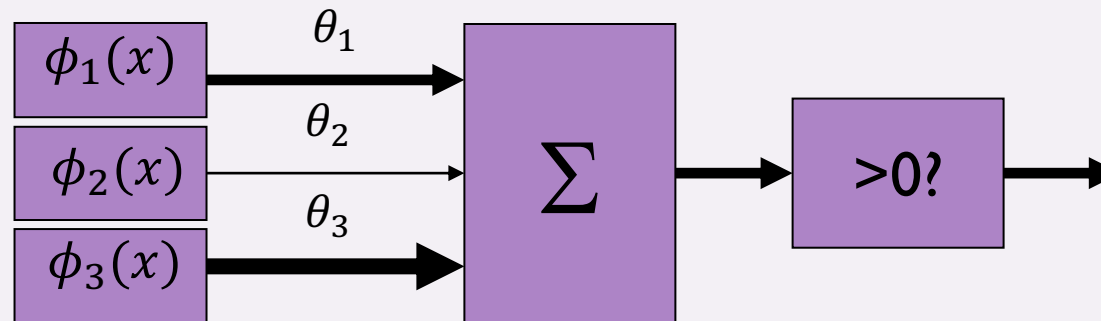
```
PIXEL-7,12 : 1  
PIXEL-7,13 : 0  
...  
NUM_LOOPS  : 1  
...
```

“2”

LINEAR CLASSIFIERS

- Inputs are **feature values**
- Each feature has a **weight**
- Weighted sum is the **activation**

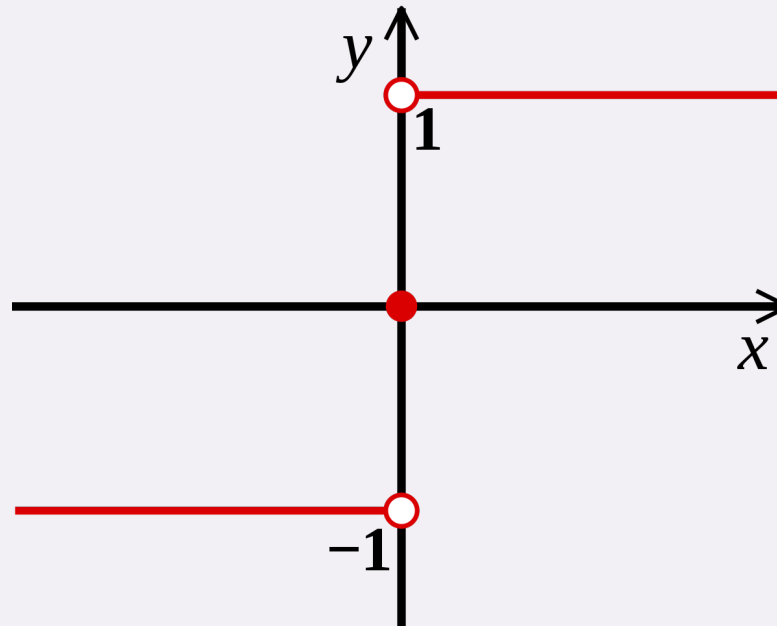
$$\text{activation}_{\theta}(x) = \sum_i \theta_i \phi_i(x) = \theta \cdot \phi(x)$$



- If the activation is:
 - Positive, output +1
 - Negative, output -1

SIGN FUNCTION

$$f_{\theta}(x) = \text{sign}(\theta \cdot \phi(x)) = \begin{cases} +1, & \text{if } \theta \cdot \phi(x) > 0 \\ -1, & \text{if } \theta \cdot \phi(x) < 0 \\ 0 \text{ or undefined,} & \text{if } \theta \cdot \phi(x) = 0 \end{cases}$$



WEIGHTS

- Binary case: compare features to a weight vector
- Learning: figure out the weight vector from examples

```
(  
  # free      : 4  
  YOUR_NAME   :-1  
  MISPELLED   : 1  
  FROM_FRIEND :-3  
  ...  
)
```

θ

$\phi(x_1)$

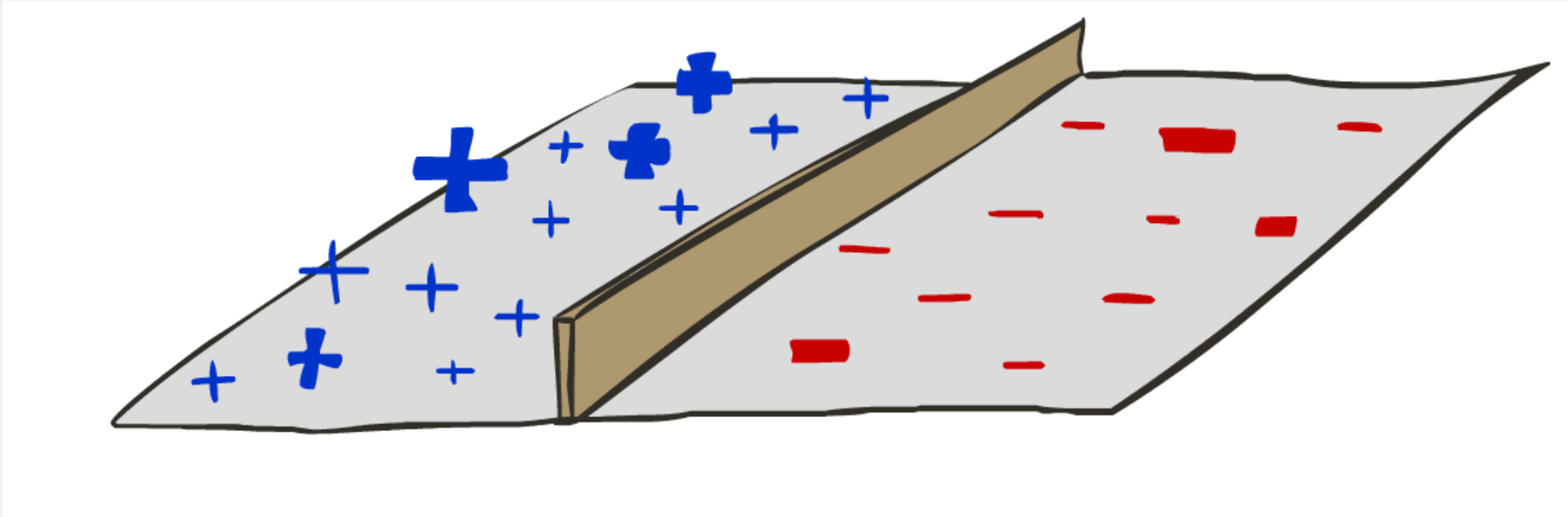
```
(  
  # free      : 2  
  YOUR_NAME   : 0  
  MISPELLED   : 2  
  FROM_FRIEND : 0  
  ...  
)
```

*Dot product $\theta \cdot \phi$ positive
means the positive class*

$\phi(x_2)$

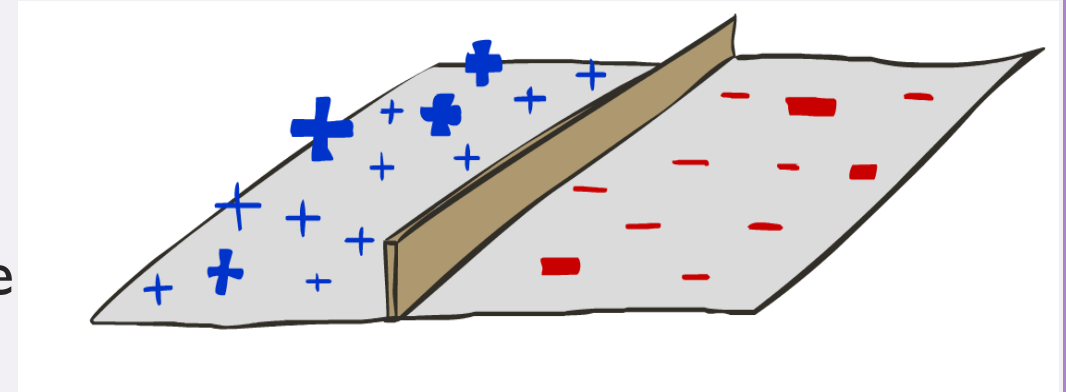
```
(  
  # free      : 0  
  YOUR_NAME   : 1  
  MISPELLED   : 1  
  FROM_FRIEND : 1  
  ...  
)
```

DECISION RULES



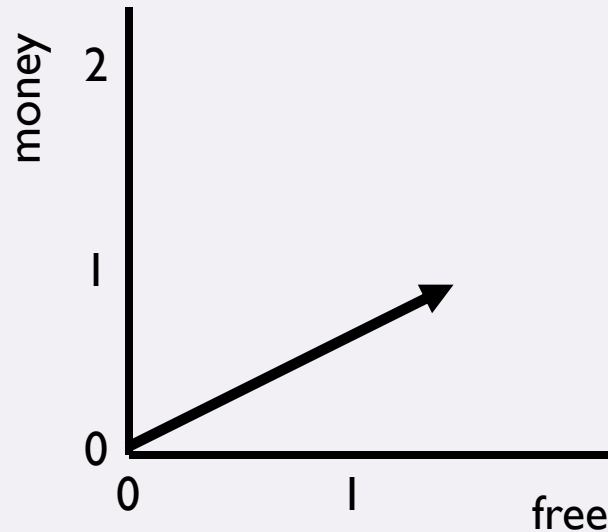
BINARY DECISION RULE

- In the space of feature vectors
 - Examples are points
 - Any weight correspond to a hyperplane
 - One side corresponds to $Y=+1$
 - Other corresponds to $Y=-1$



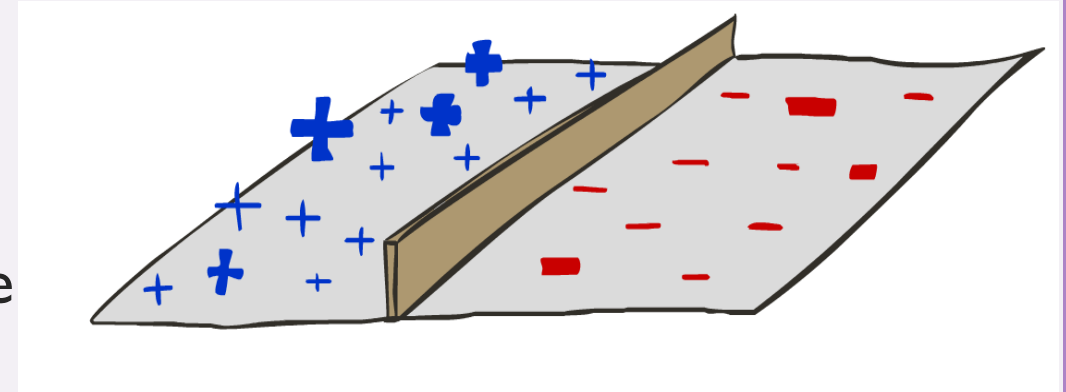
θ

BIAS	:	-4
free	:	4
money	:	2
...		



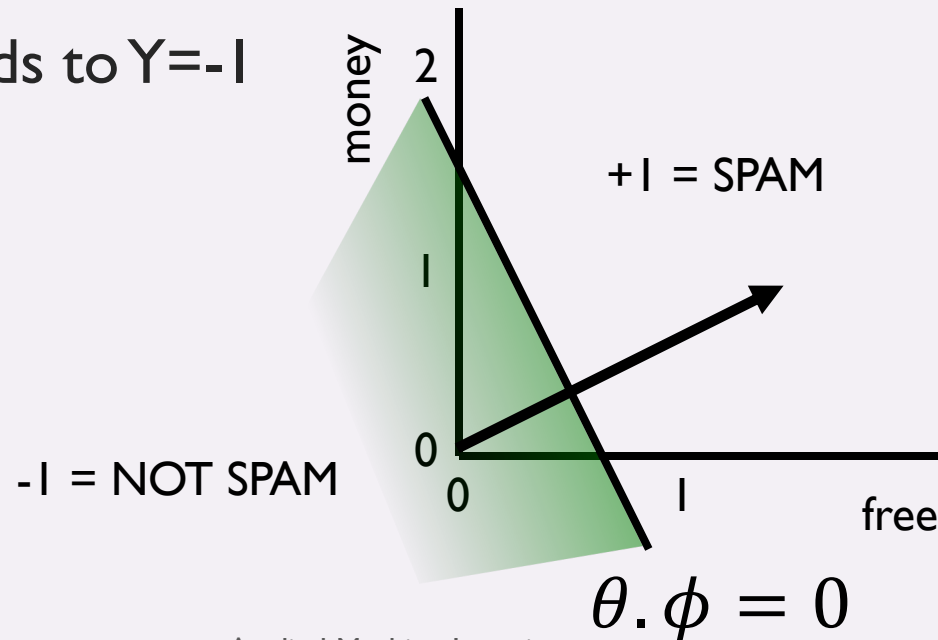
BINARY DECISION RULE

- In the space of feature vectors
 - Examples are points
 - Any weight correspond to a hyperplane
 - One side corresponds to $Y=+1$
 - Other corresponds to $Y=-1$

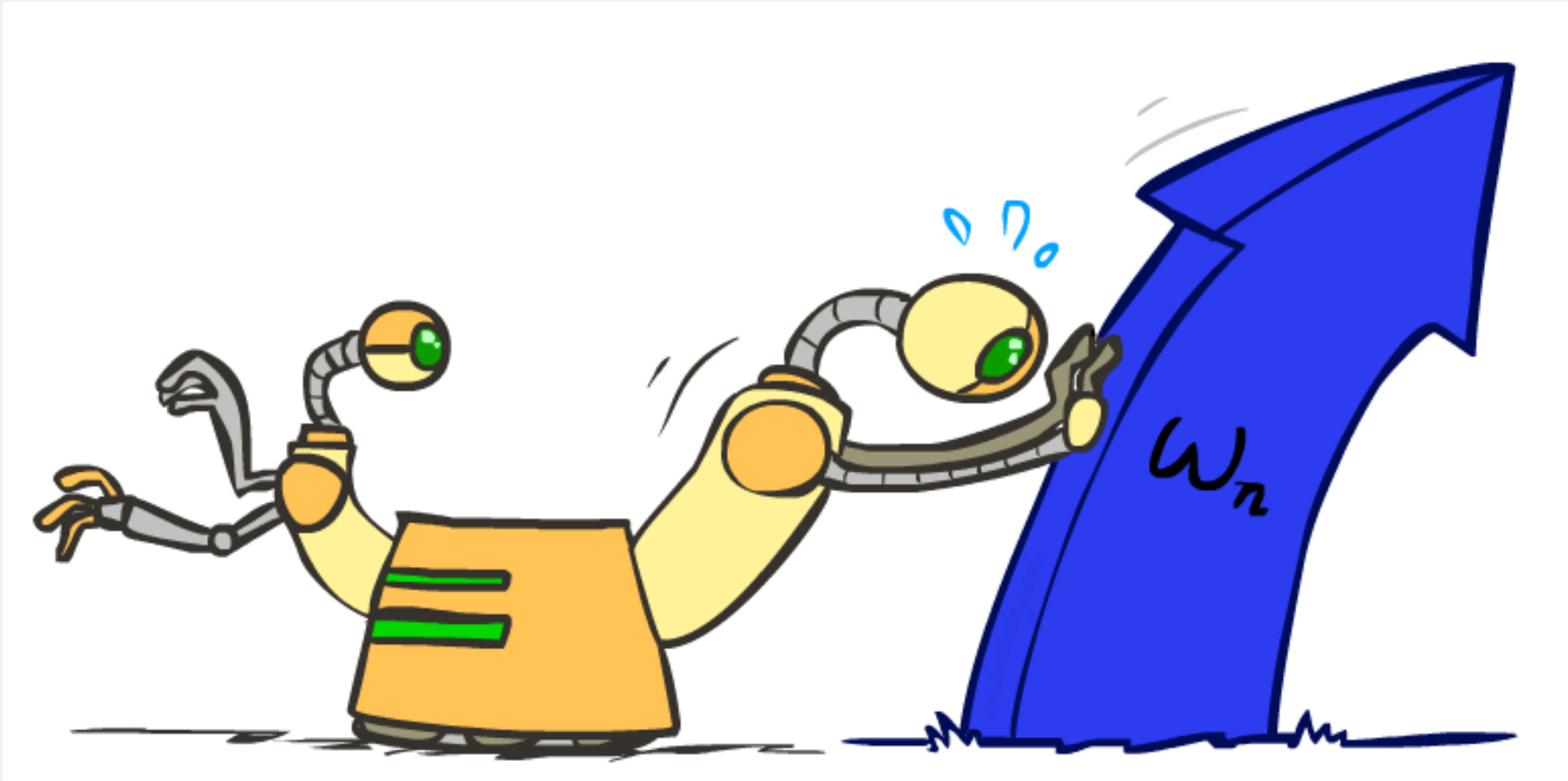


θ

BIAS	:	-4
free	:	4
money	:	2
...		

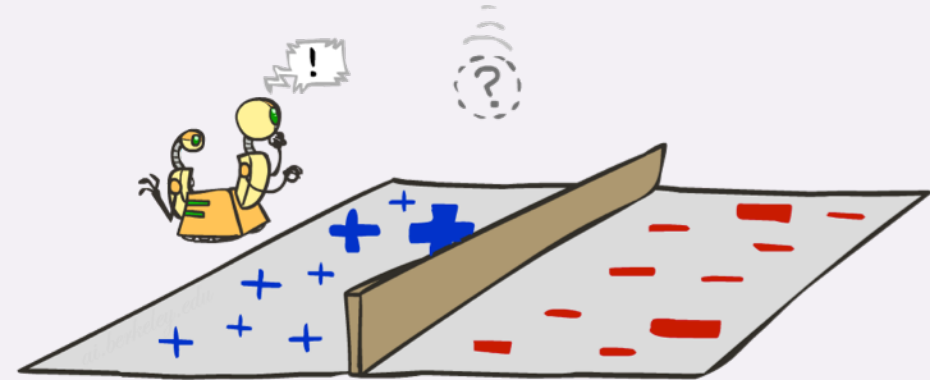


WEIGHT UPDATES

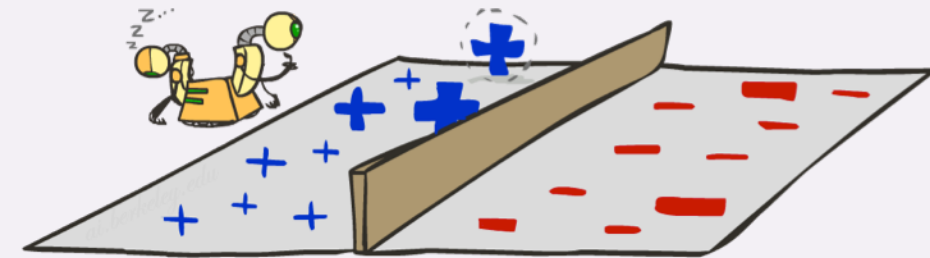


LEARNING: BINARY CLASSIFIER

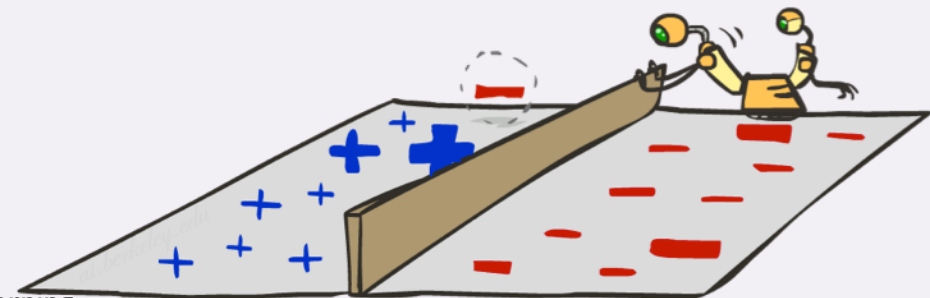
- Start with weights = 0
- For each training instance:
 - Classify with current weights



- If correct (i.e., $Y=Y^*$), no change!

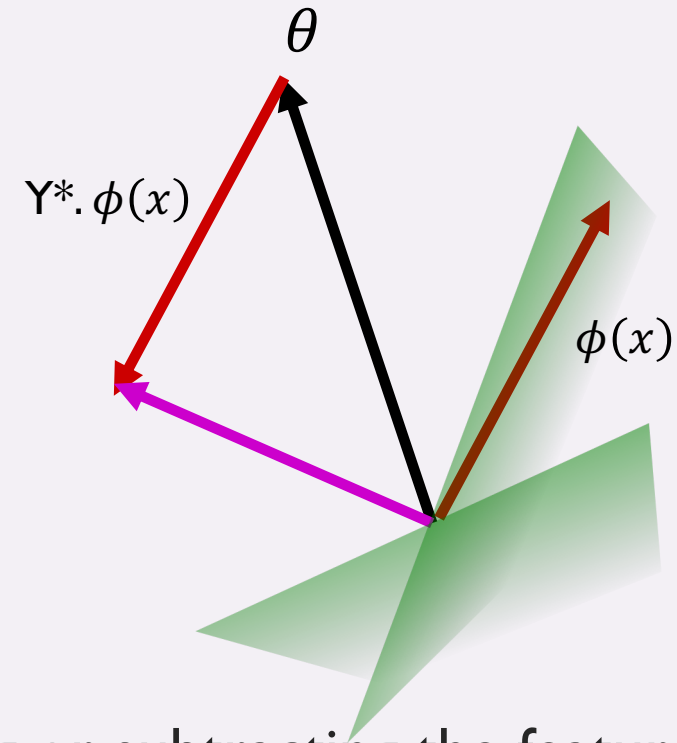


- If wrong: adjust the weight vector



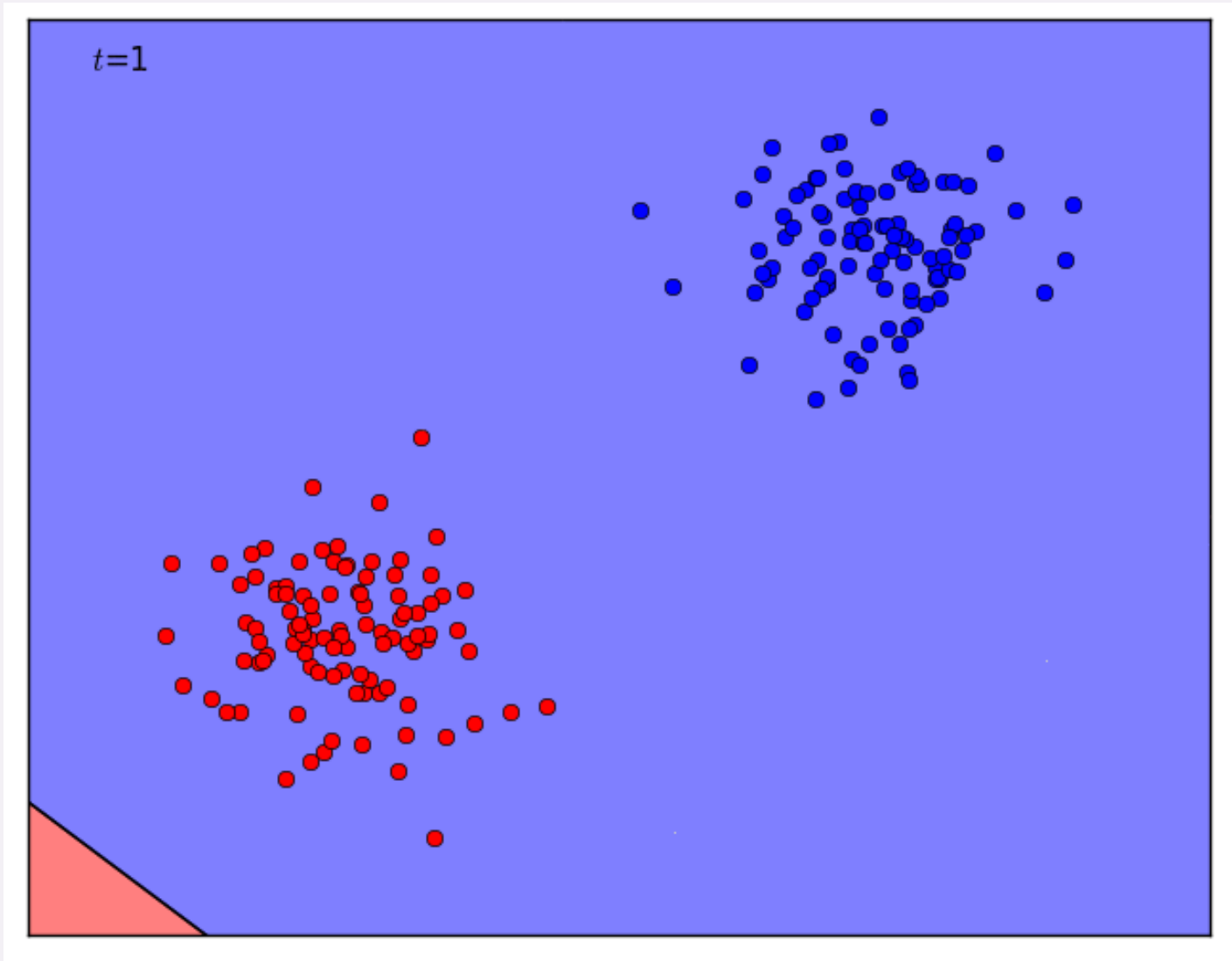
LEARNING: BINARY CLASSIFIER

- Start with weights = 0
- For each training instance:
 - Classify with current weights
 - If correct (i.e., $Y=Y^*$), no change!
 - If wrong: adjust the weight vector by adding or subtracting the feature vector.



$$\theta = \theta + Y^* \cdot \phi(x)$$

EXAMPLE



MULTICLASS DECISION RULE

- If we have multiple classes:
 - A weight vector for each class:

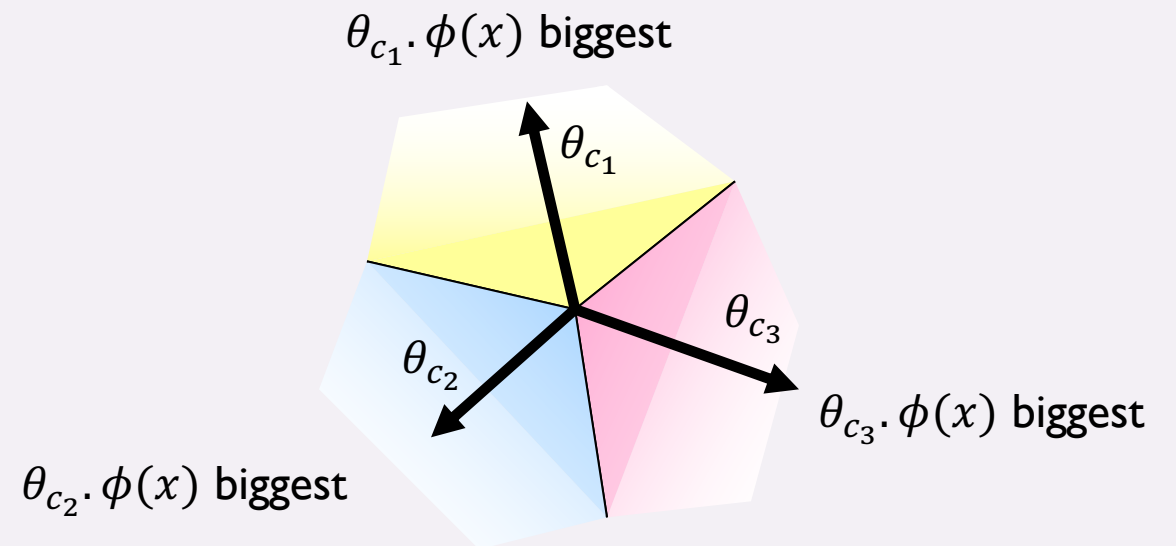
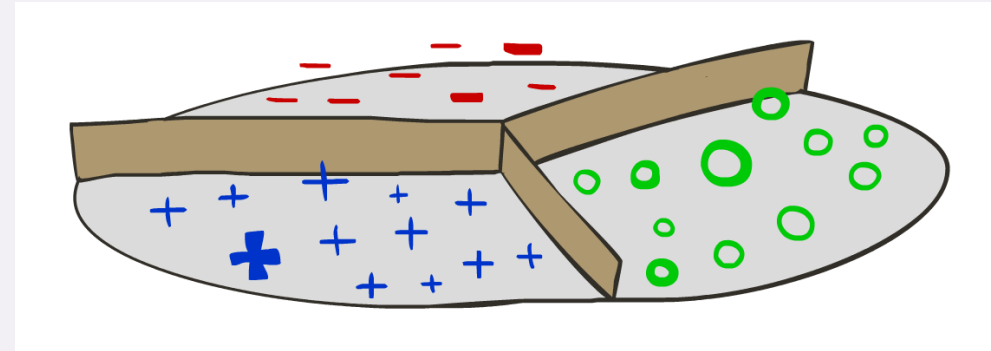
$$\theta_c$$

- Score (activation) of a class c :

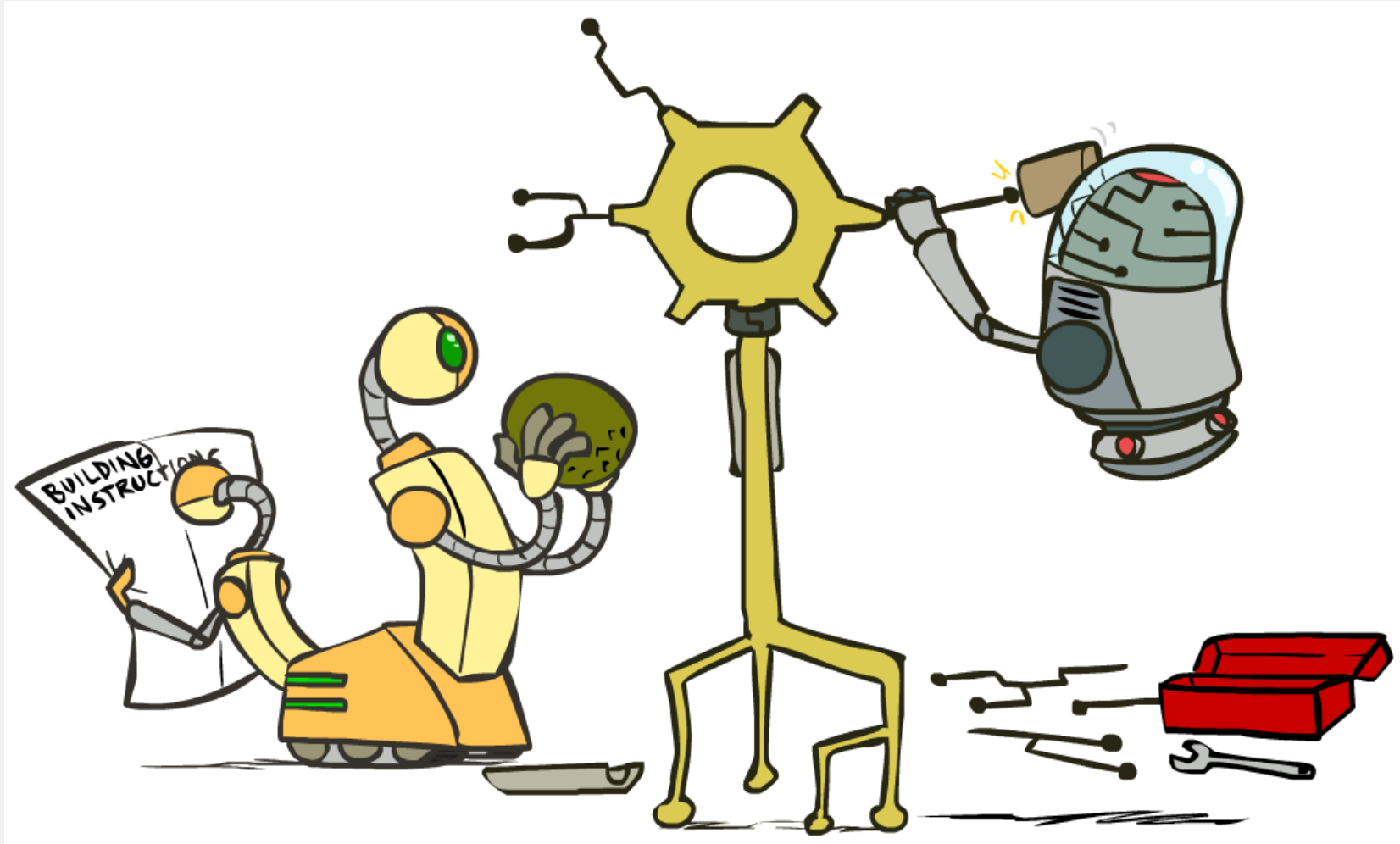
$$\theta_c \cdot \phi(x)$$

- Prediction highest score wins

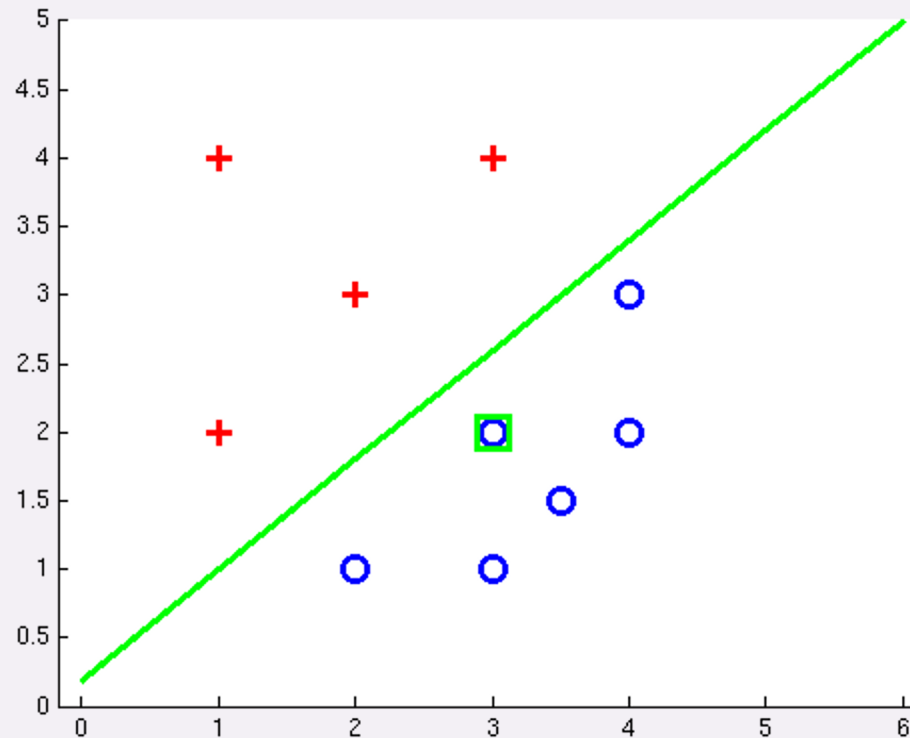
$$y = \underset{c}{\operatorname{argmax}} \theta_c \cdot \phi(x)$$



IMPROVING THE CLASSIFIER

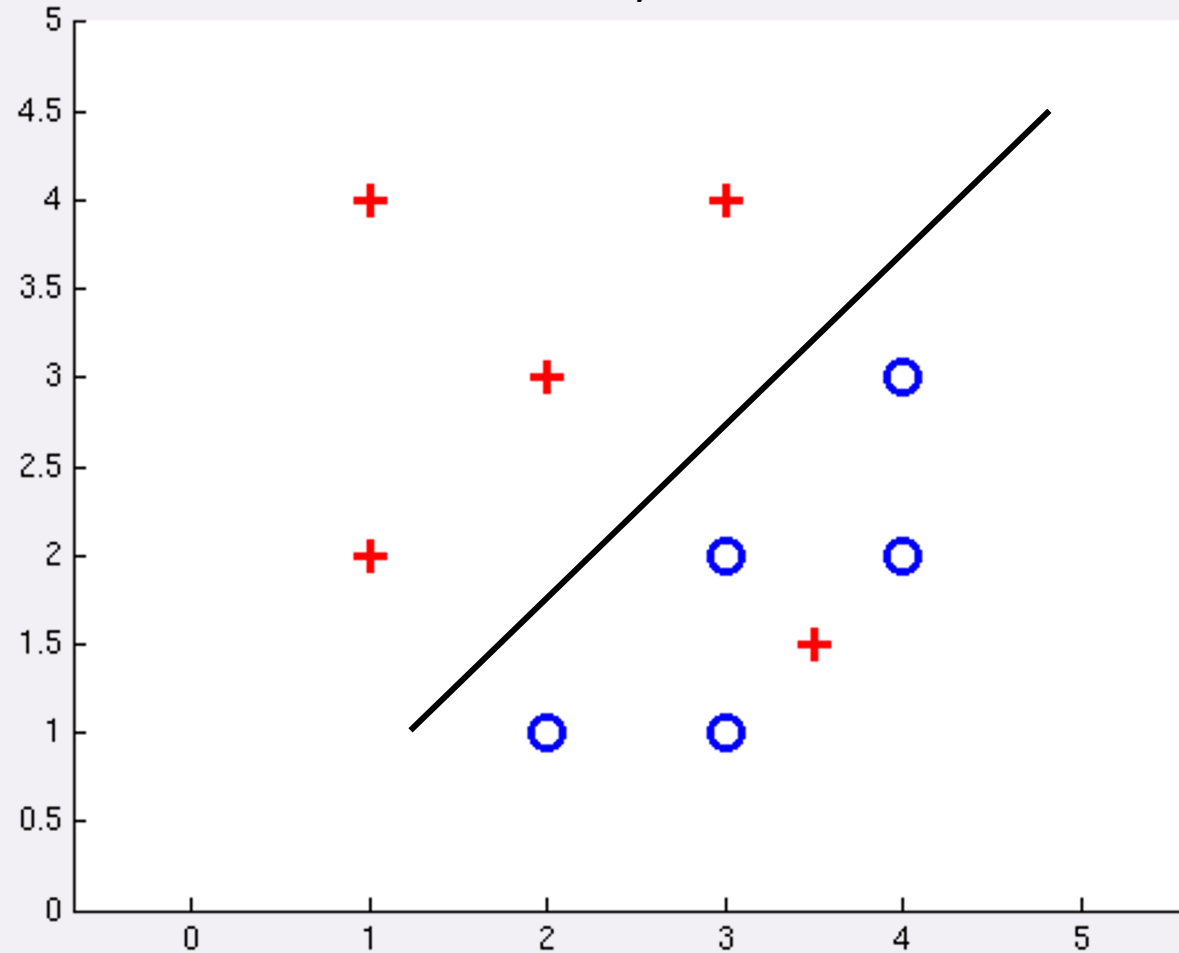


SEPARABLE CASE: DETERMINISTIC DECISION

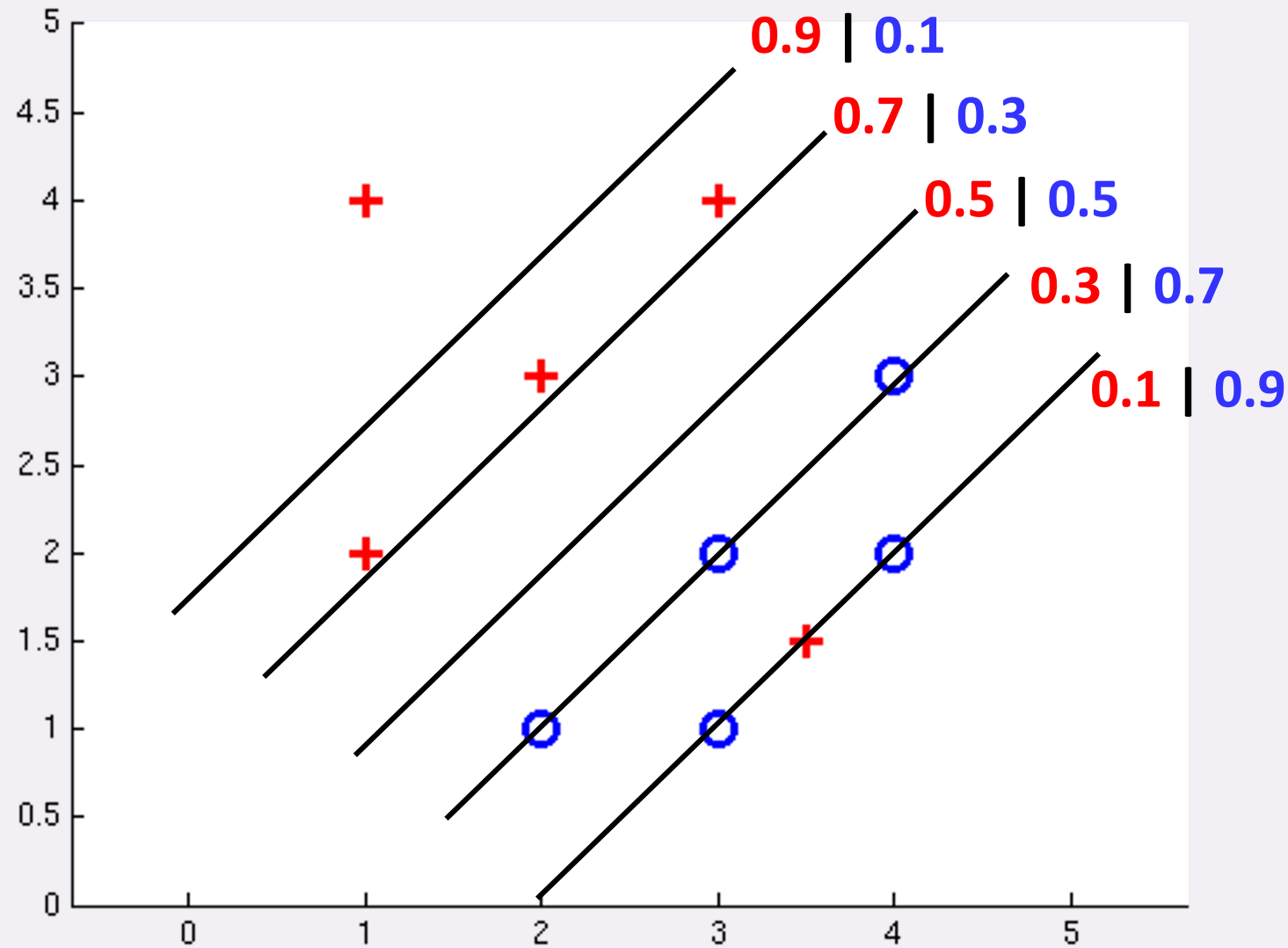


NON-SEPARABLE CASE: DETERMINISTIC DECISION

Even the best linear boundary makes at least one mistake



PROBABILISTIC DECISION



HW1 (DUE TONIGHT)



HW 1

Due Apr 20 at 11:59pm | 60 pts



QUIZ 1 (DUE THURSDAY)



Quiz 2

Not available until Apr 22 at 3:00pm | Due Apr 22 at 11:59pm | 7 pts | 7 Questions



A decorative graphic on the left side of the slide consisting of two parallel, wavy vertical lines. The inner line is a light purple color, and the outer line is a slightly darker shade of purple. They extend from the top to the bottom of the slide.

QUESTIONS?