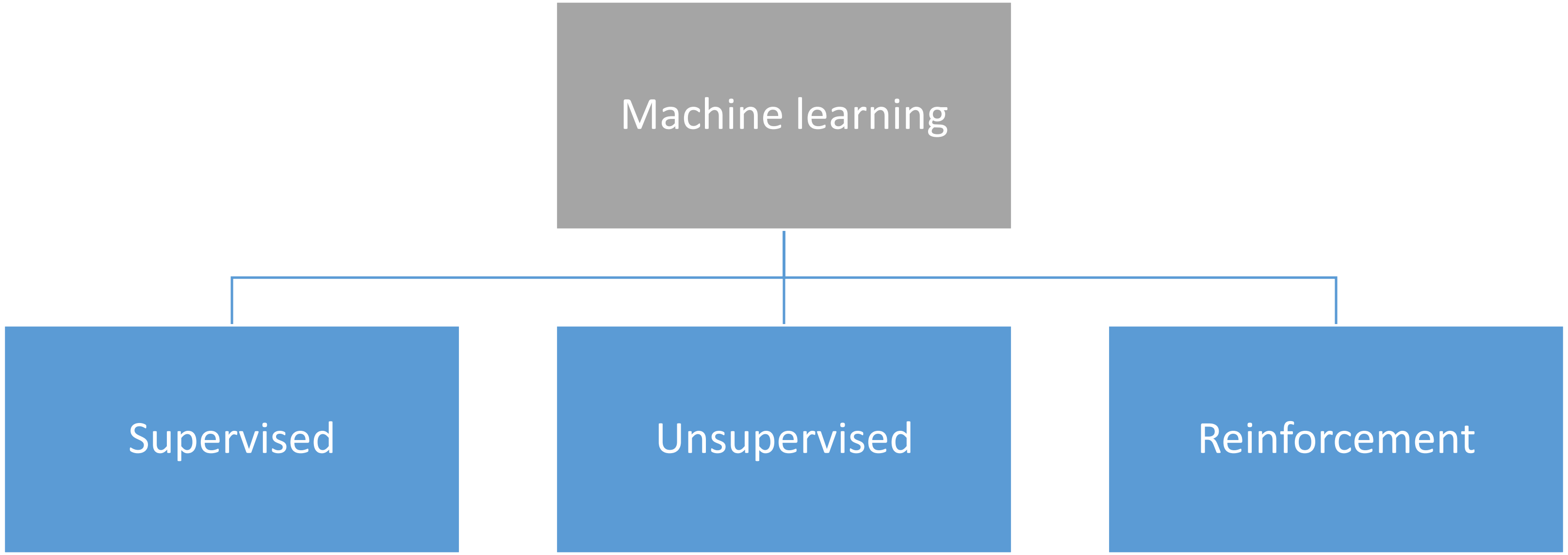
Machine Learning

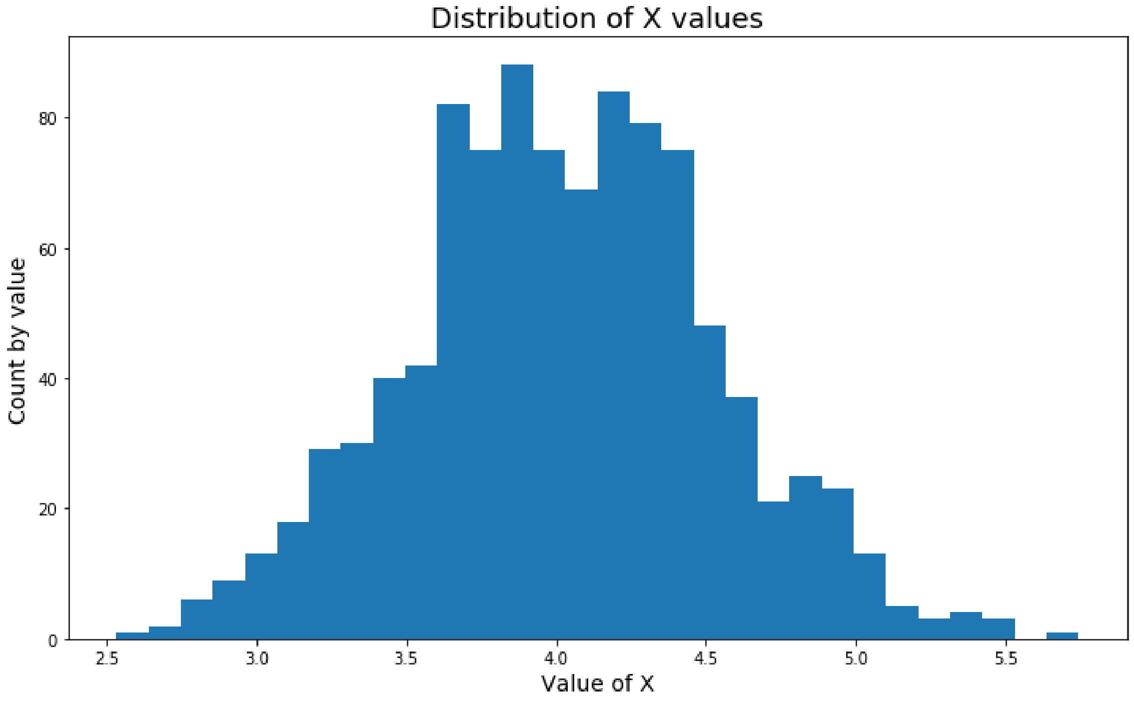
Getting There

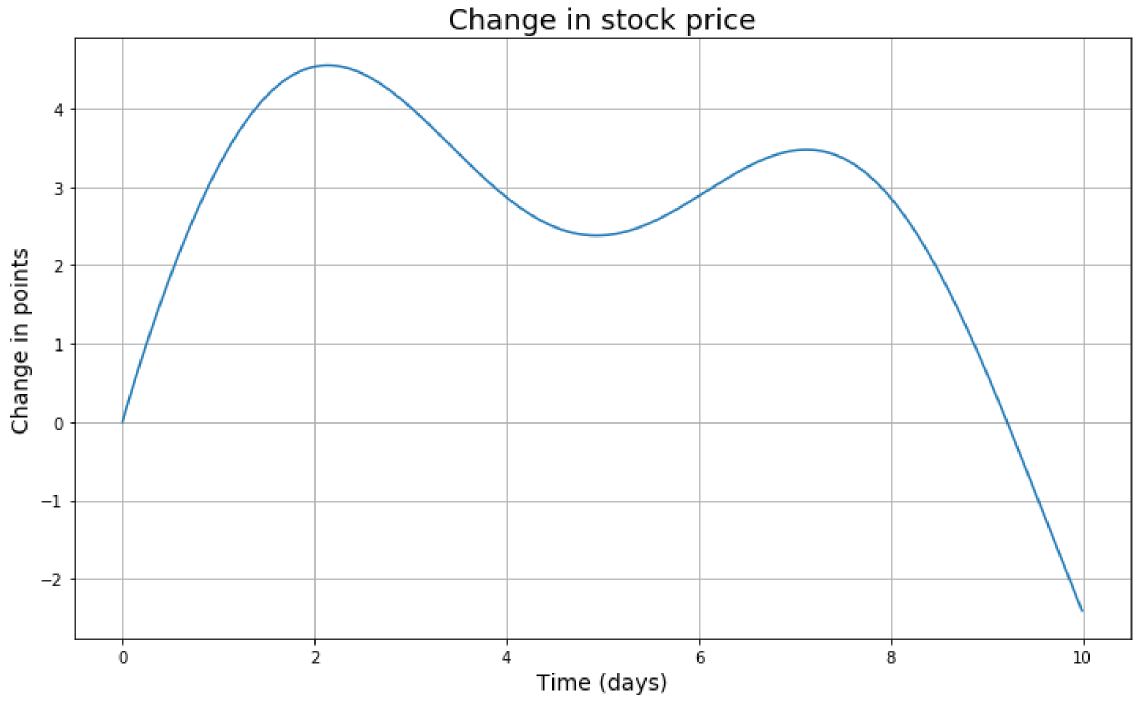
Types of machine learning



NYC taxi dataset

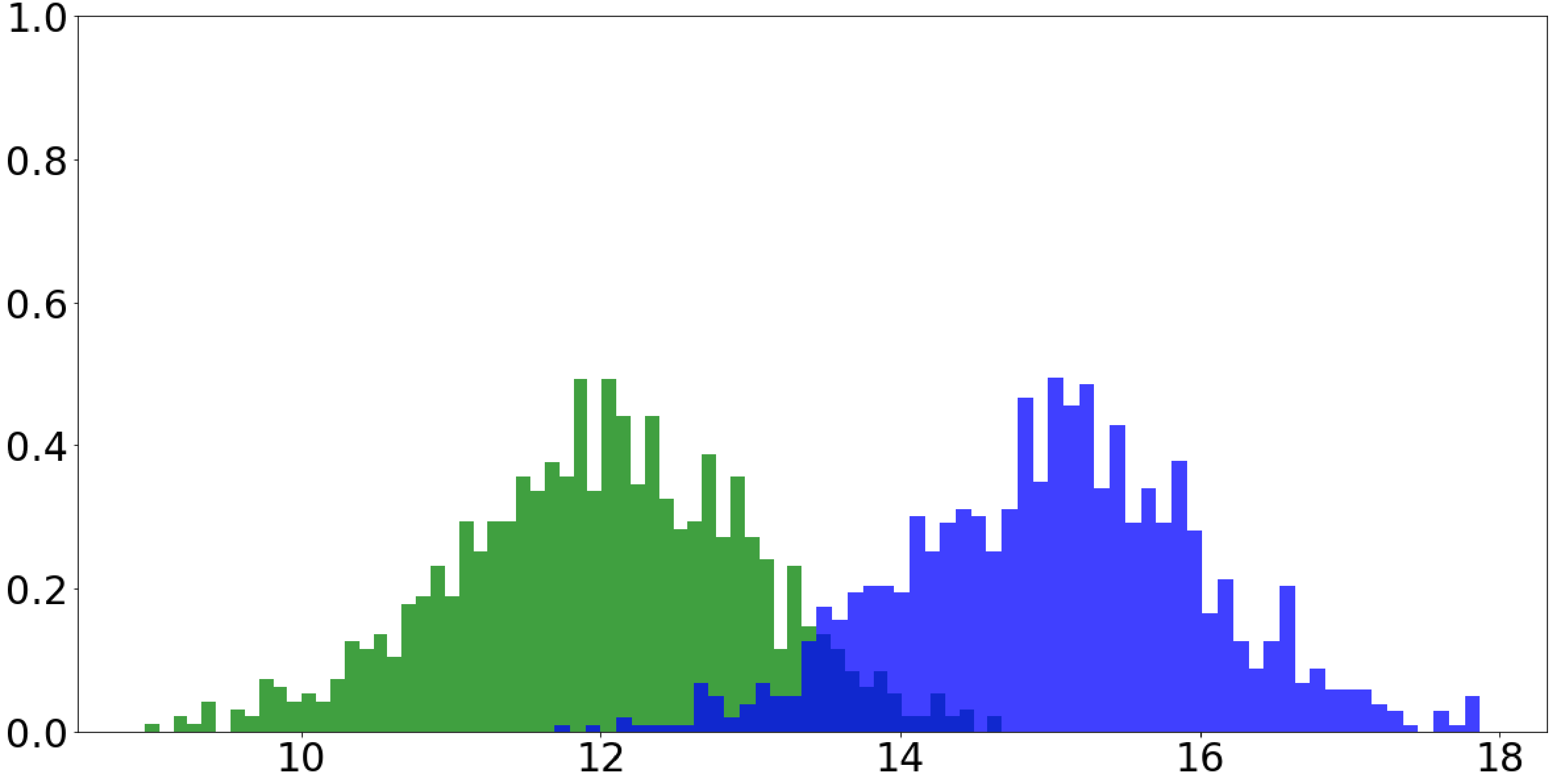
http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

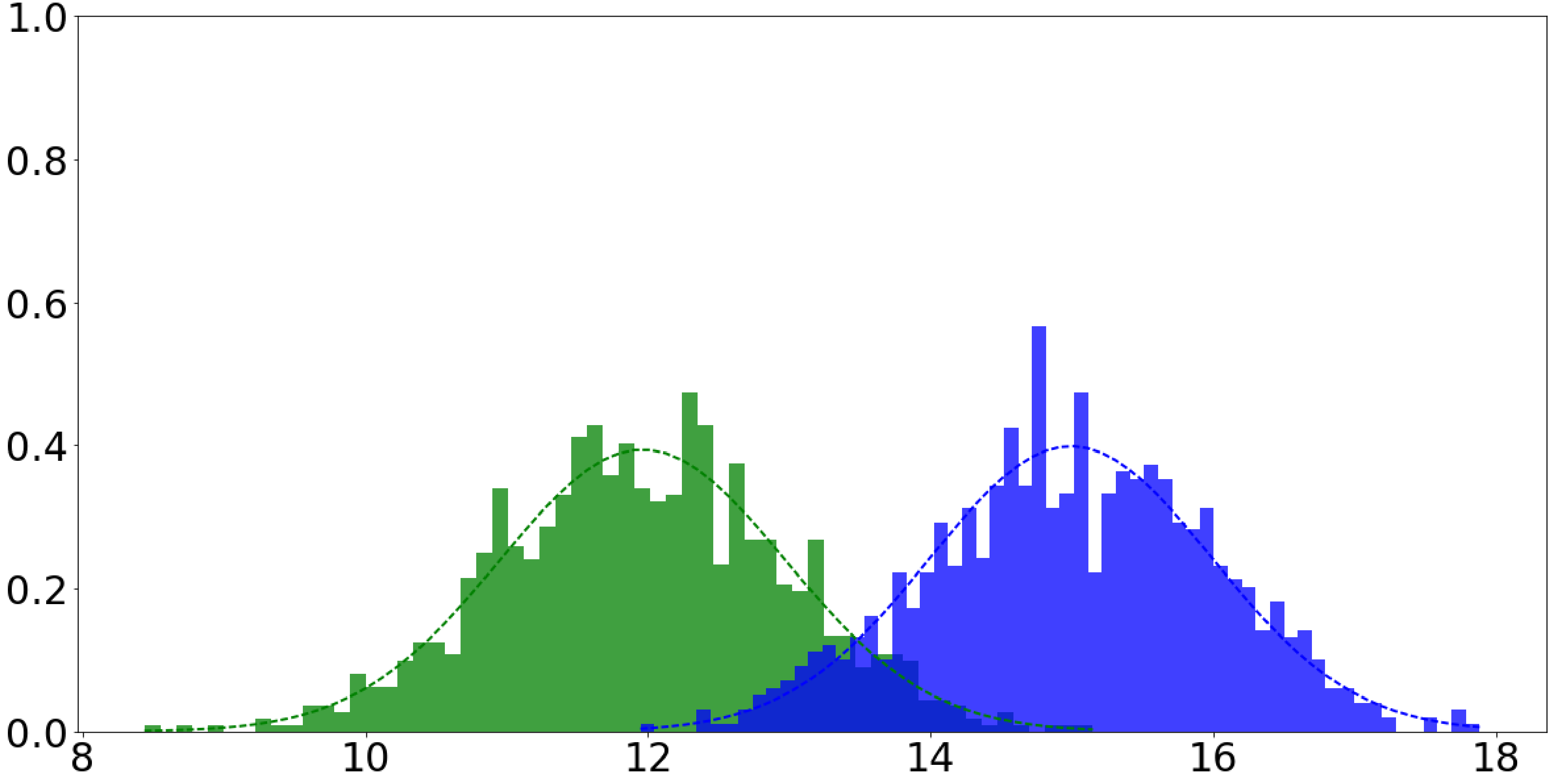




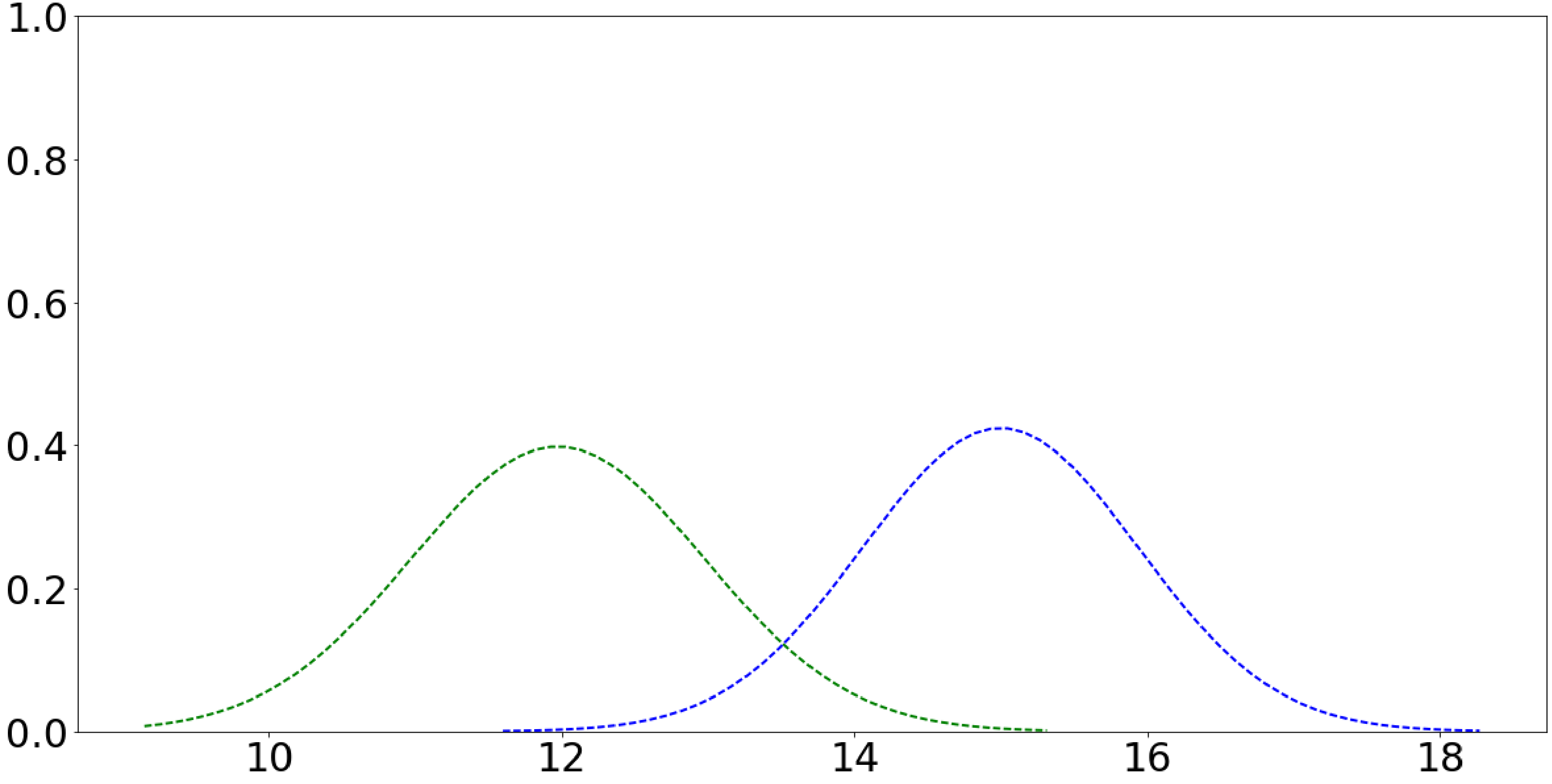
Visualization

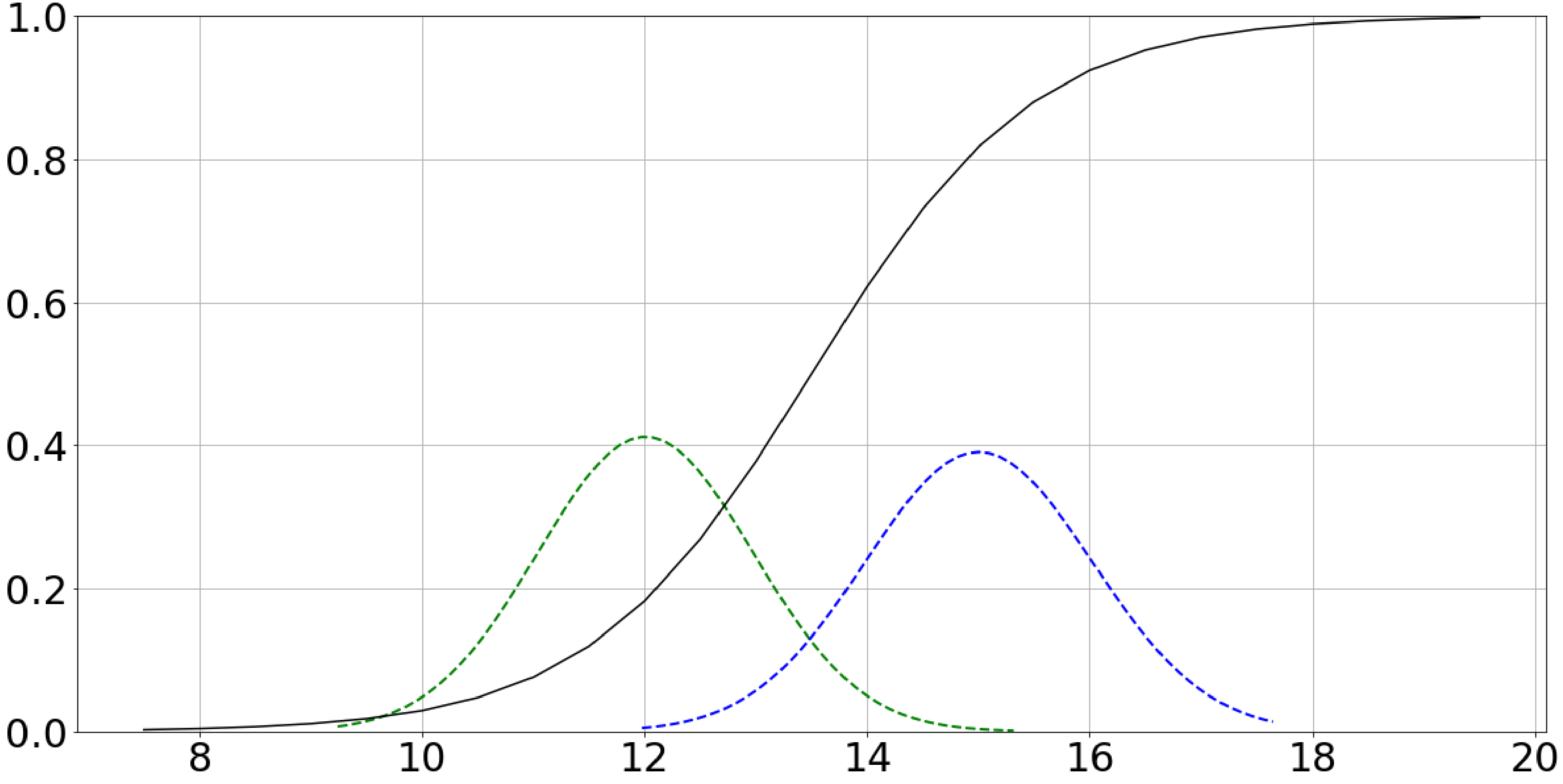
Classification

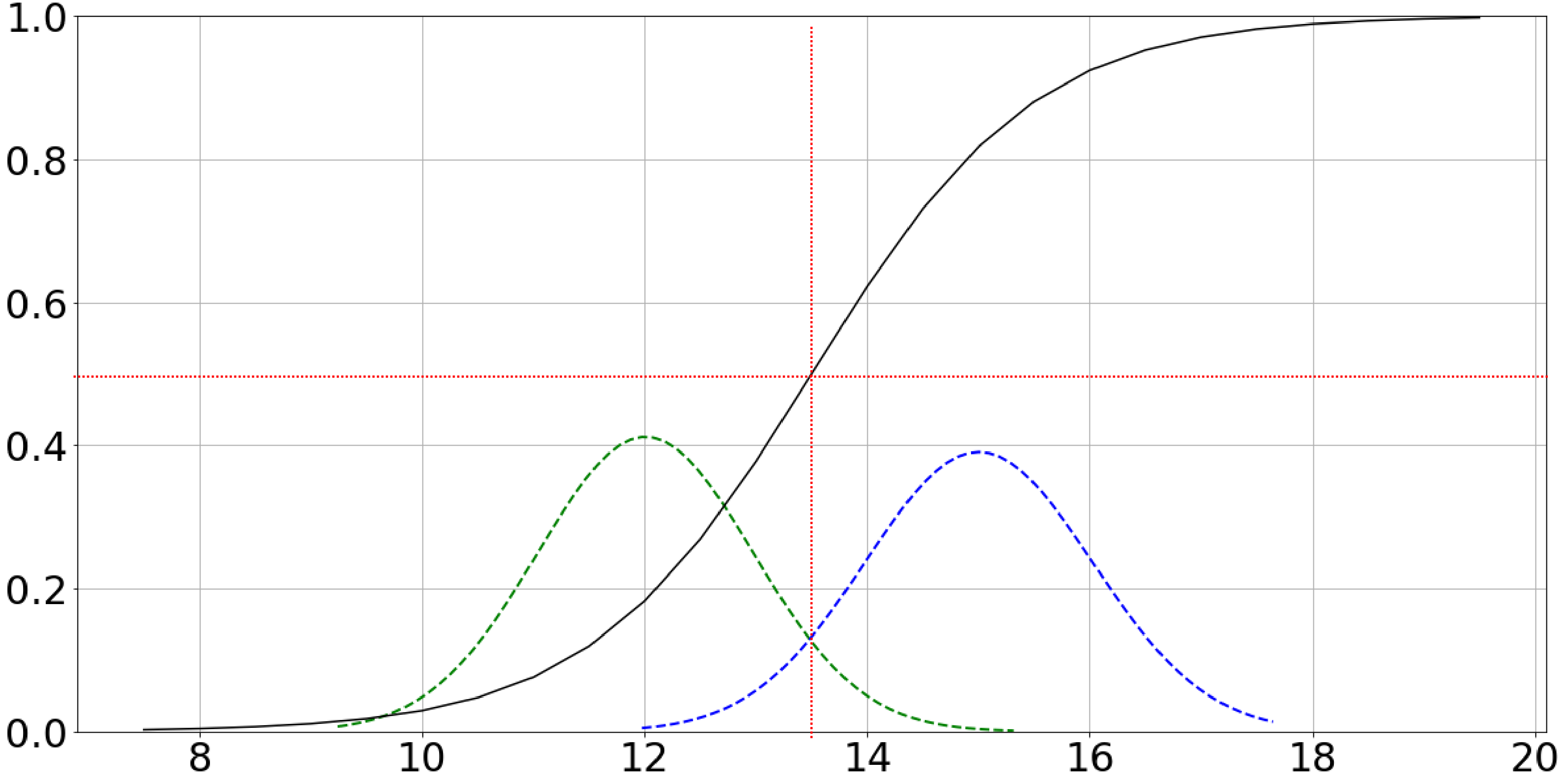

1.0
0.8
0.6
0.4
0.2
0.0
10
12
14
16
18

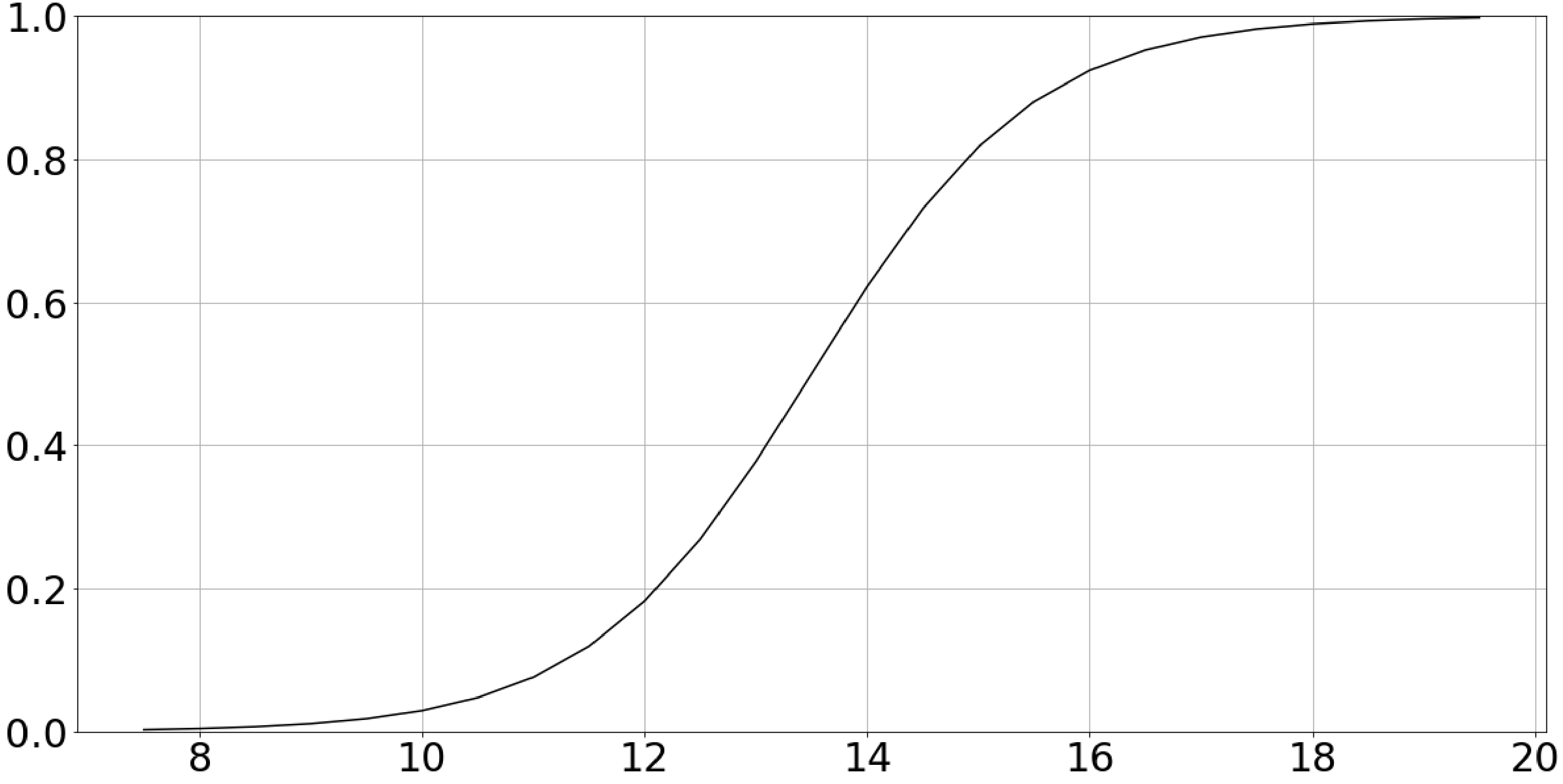


1.0
0.8
0.6
0.4
0.2
0.0
8
10
12
14
16
18


1.0
0.8
0.6
0.4
0.2
0.0
10
12
14
16
18

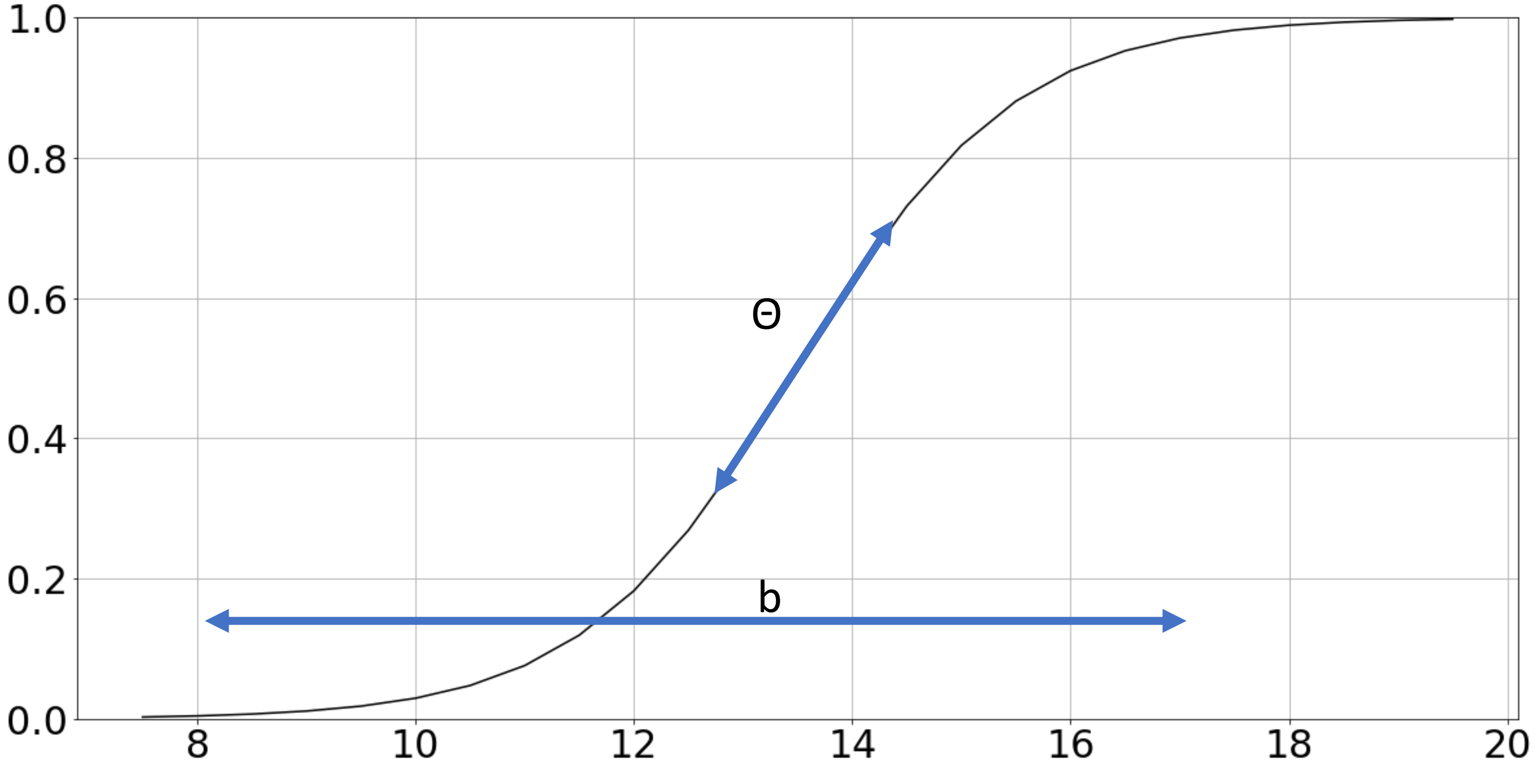

1.0
0.8
0.6
0.4
0.2
0.0
8
10
12
14
16
18
20

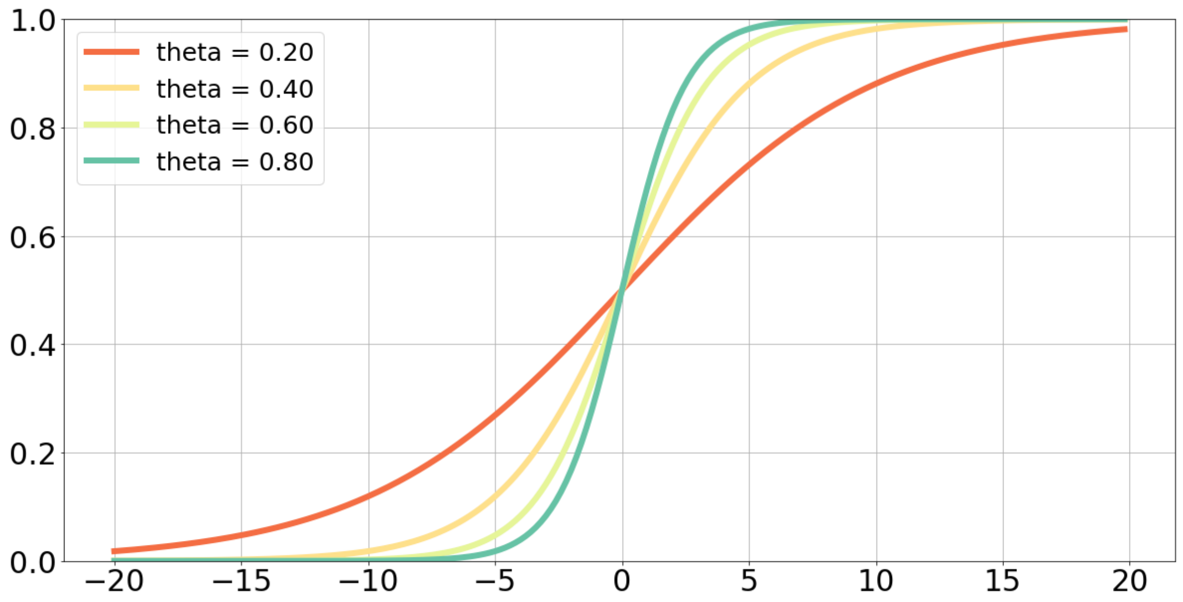

1.0
0.8
0.6
0.4
0.2
0.0
8
10
12
14
16
18
20

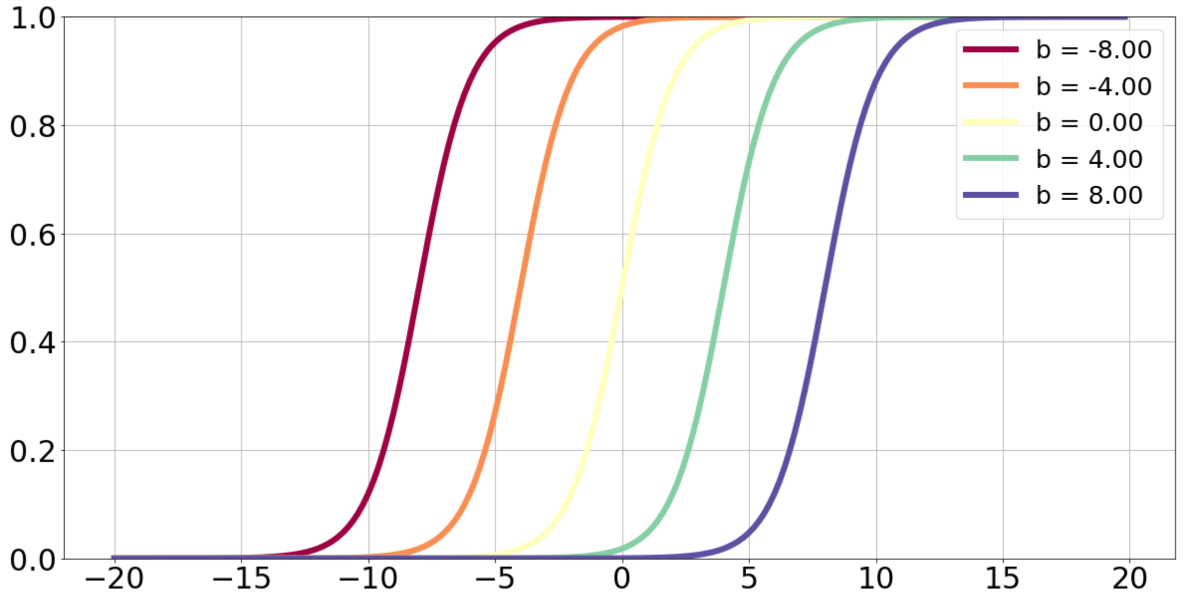

1.0
0.8
0.6
0.4
0.2
0.0
8
10
12
14
16
18
20

The sigmoid function

$$y = sigmoid(X\theta\ - b) = \frac{1}{1 + e^{-X\theta + b}}$$

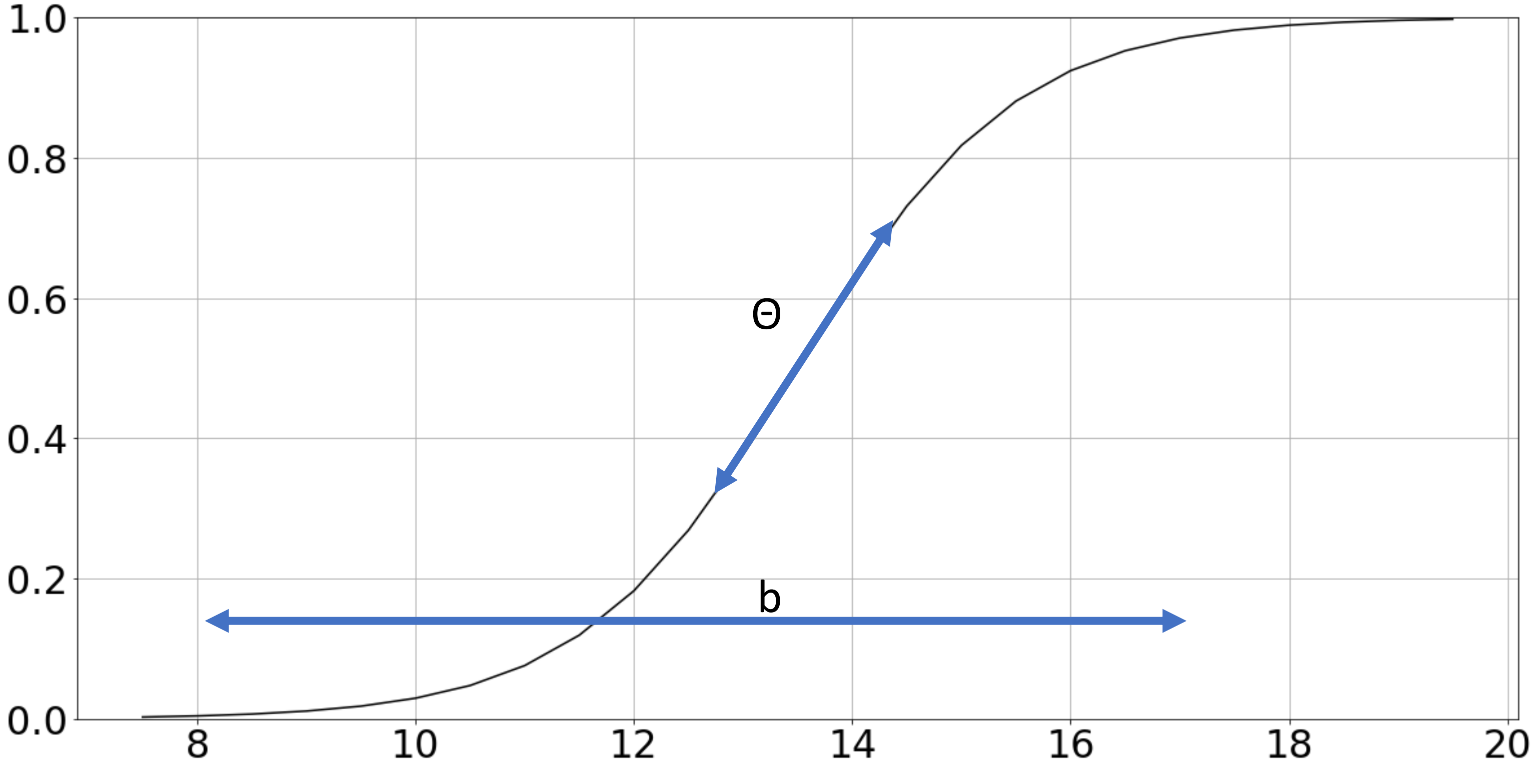



theta = 0.20
theta = 0.40
theta = 0.60
theta = 0.80
1.0
0.8
0.6
0.4
0.2
0.0
−20
−15
−10
−5
0
5
10
15
20

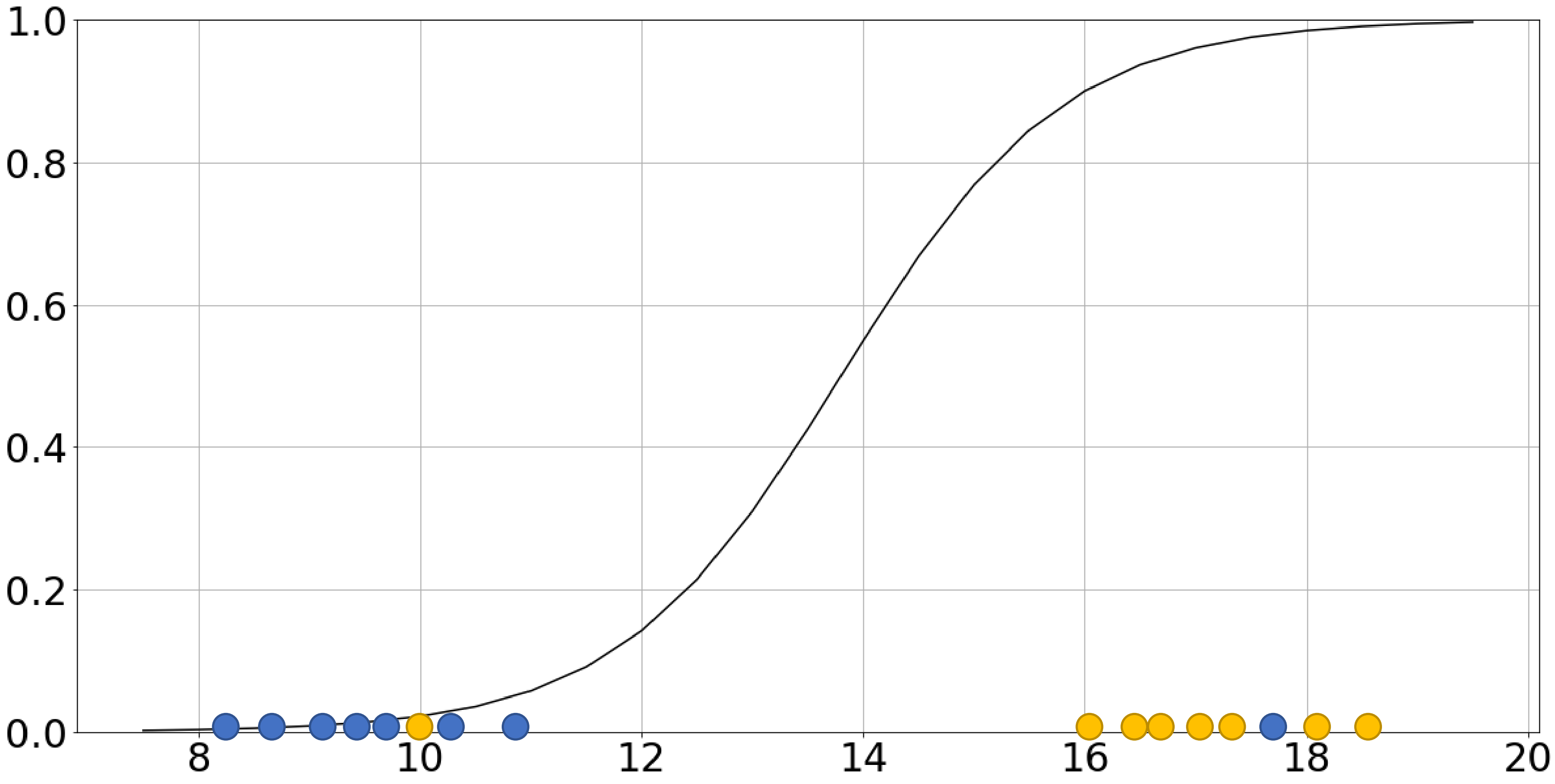

b = -8.00
b = -4.00
b = 0.00
b = 4.00
b = 8.00
1.0
0.8
0.6
0.4
0.2
0.0
−20
−15
−10
−5
0
5
10
15
20

The sigmoid function

$$y = sigmoid(X\theta \ - b) = \frac{1}{1 + e^{-X\theta + b}}$$

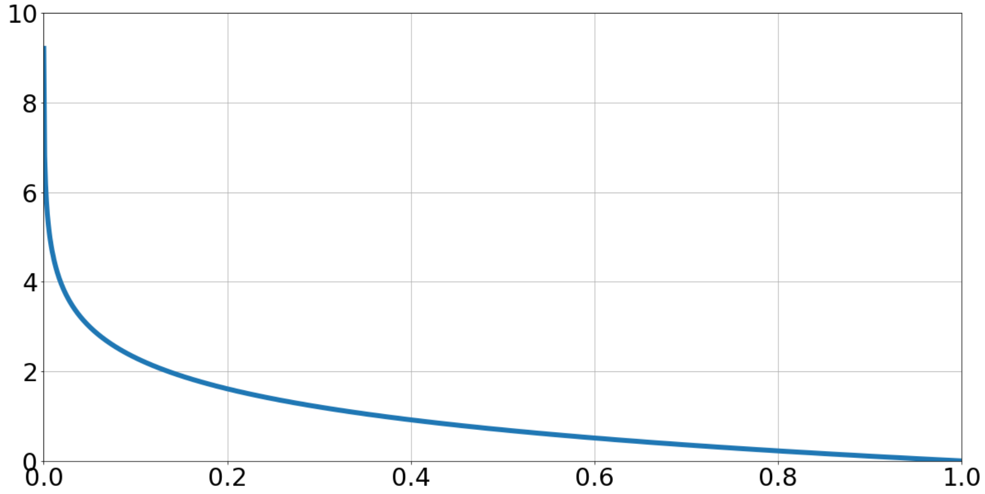


Loss function for the logistic regression



Loss function for logistic regression

$$c(\theta|x) = \begin{cases} -\log\big(h(x;\theta)\big) & y = 1 \\ -\log\big(1 - h(x;\theta)\big) & y = 0 \end{cases}$$



$$h(x;\theta) = \frac{1}{1 + e^{-X\theta + b}}$$

Loss function for logistic regression

$$c(\theta) = \begin{cases} -\log(h(x)), & y = 1 \\ -\log(1 - h(x)), & y = 0 \end{cases}$$

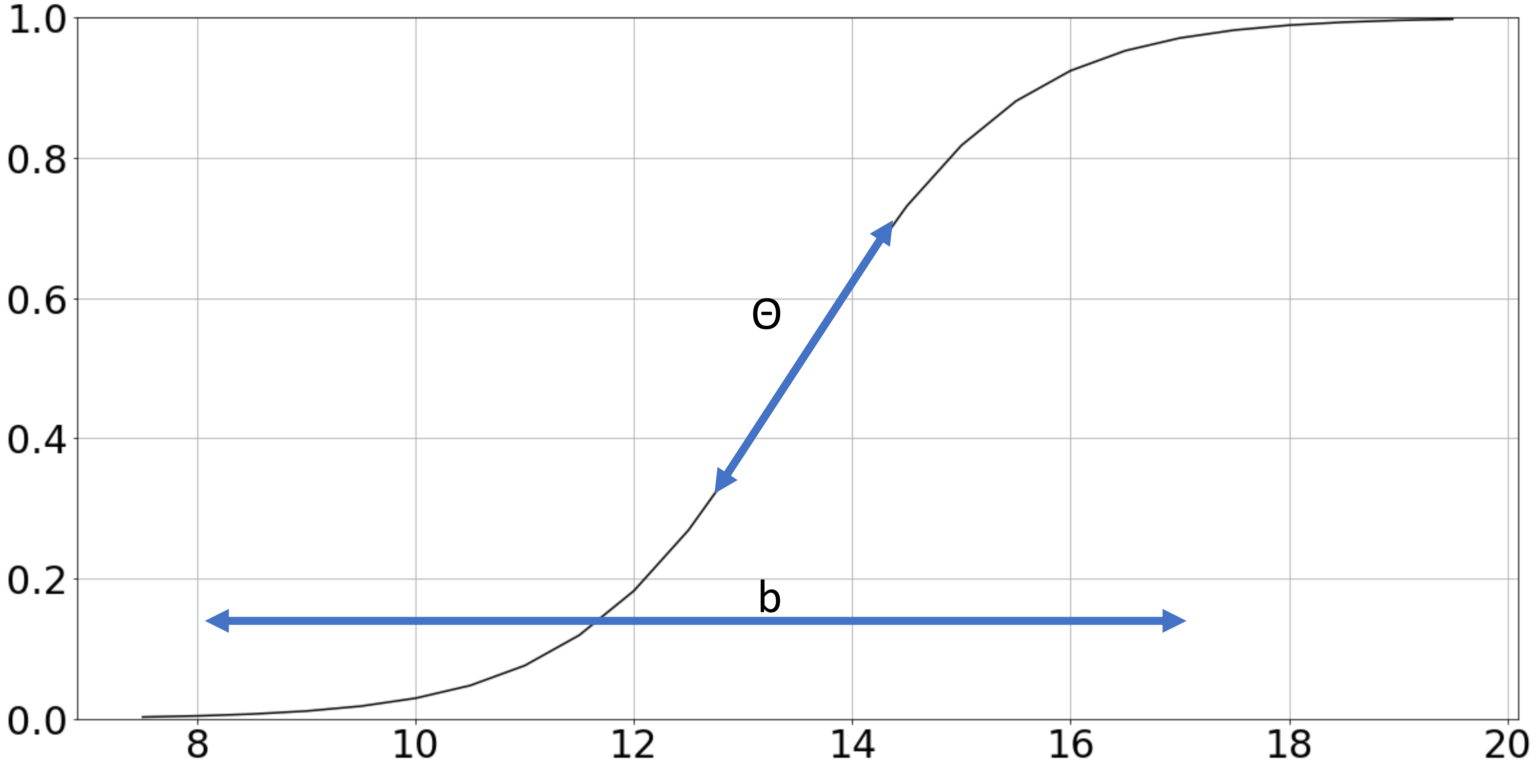
$$c(\theta) = -\log\big(h(x)\big)^{y} - \log\big(1 - h(x)\big)^{1-y}$$

$$c(\theta) = -[\tfrac{1}{m}\textstyle\sum_{i=0}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)]$$

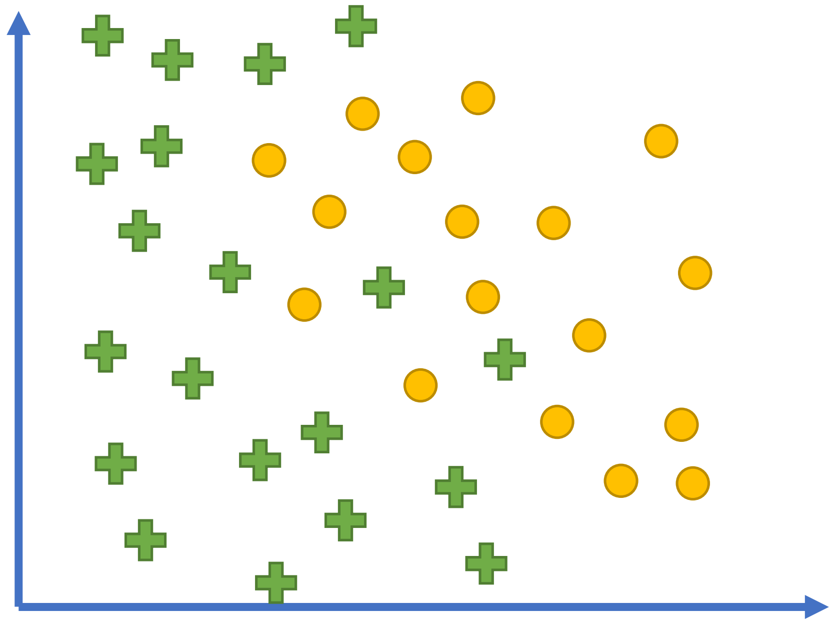
$$\frac{\partial c(\theta)}{\partial \theta_j} = \frac{\alpha}{m} \sum_i \left(h\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

The sigmoid function

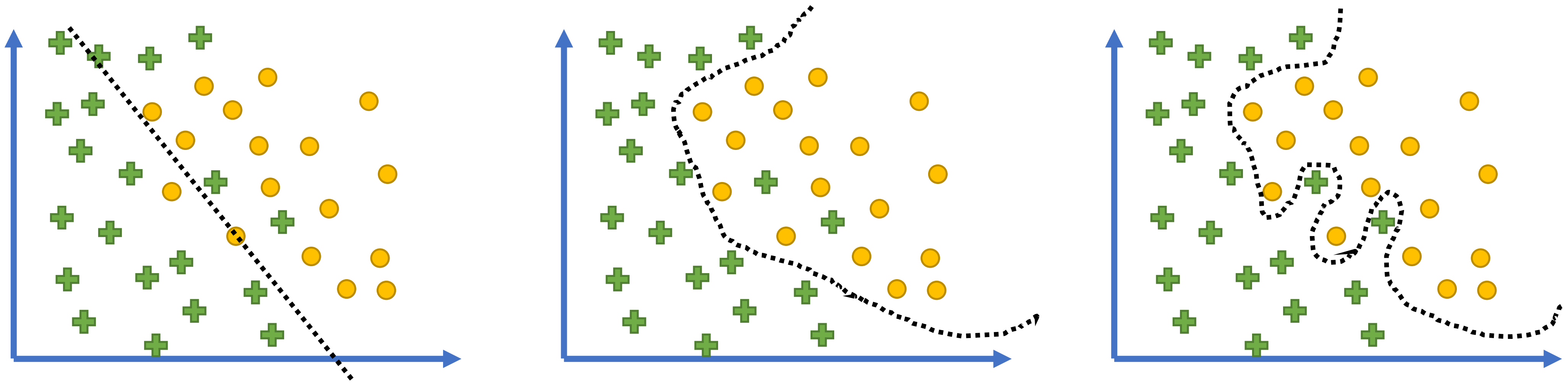
$$y = sigmoid(X\theta \ - b) = \frac{1}{1 + e^{-X\theta + b}}$$



Bias vs Variance



Bias vs Variance



Bias vs Variance

- How balance **overfitting** and **expressiveness**?
- How to **estimate this tradeoff**?

Generalization problem

Model with an accuracy of 99.9%

99.9%

$$P[|E(X) - \mu(X)| > \epsilon] \leq 2\mathrm{e}^{-2\epsilon^2 N}$$

https://www.flickr.com/photos/calliope/534526381
https://upload.wikimedia.org/wikipedia/commons/thumb/1280/gpj.01_emarugaK_ekaL_ni_sesirI/7/75px-Irises_in_Lake_Kagurame_gpj.01

Generalization problem

$$P[|E(X) - \mu(X)| > \epsilon] \leq 2\mathrm{e}^{-2\epsilon^2 N}$$

Model with an accuracy of 99%

Model with an accuracy of 98%

Model with an accuracy of 10%

Model with an accuracy of 14%

Model with an accuracy of 86%

.

.

.

Generalization problem

$$P[|E(X) - \mu(X)| > \epsilon] \leq 2\mathrm{e}^{-2\epsilon^2 N}$$

98.4%

train

test

97.5%

Model with an accuracy of 99%

Model with an accuracy of 98%

Model with an accuracy of 10%

Model with an accuracy of 14%

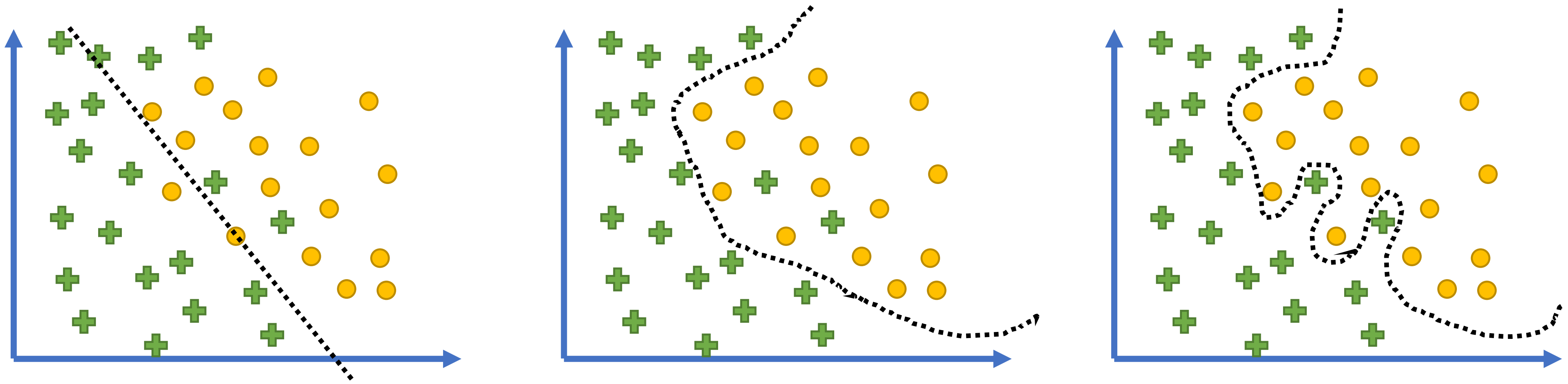
Model with an accuracy of 86%

.

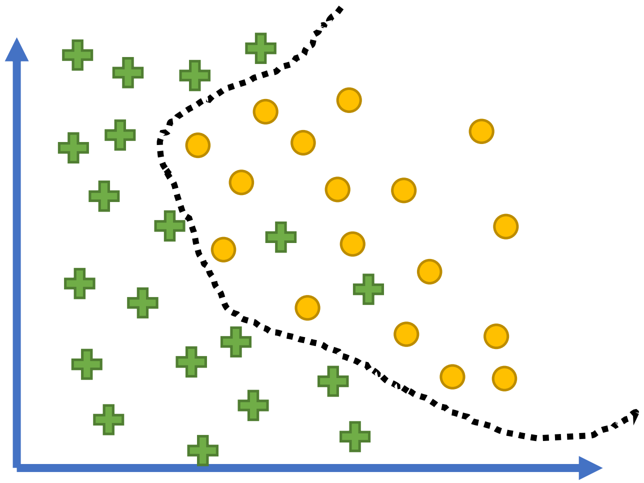
.

.

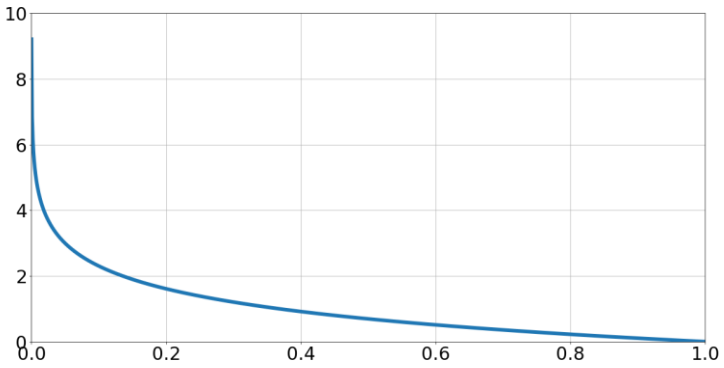
Bias vs Variance



Regularization



$$c(\theta) = -\left[\frac{1}{m}\sum_{i=0}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)\right]$$



$$c(\theta) = -\left[\frac{1}{m}\sum_{i=0}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2}||\theta||^2$$