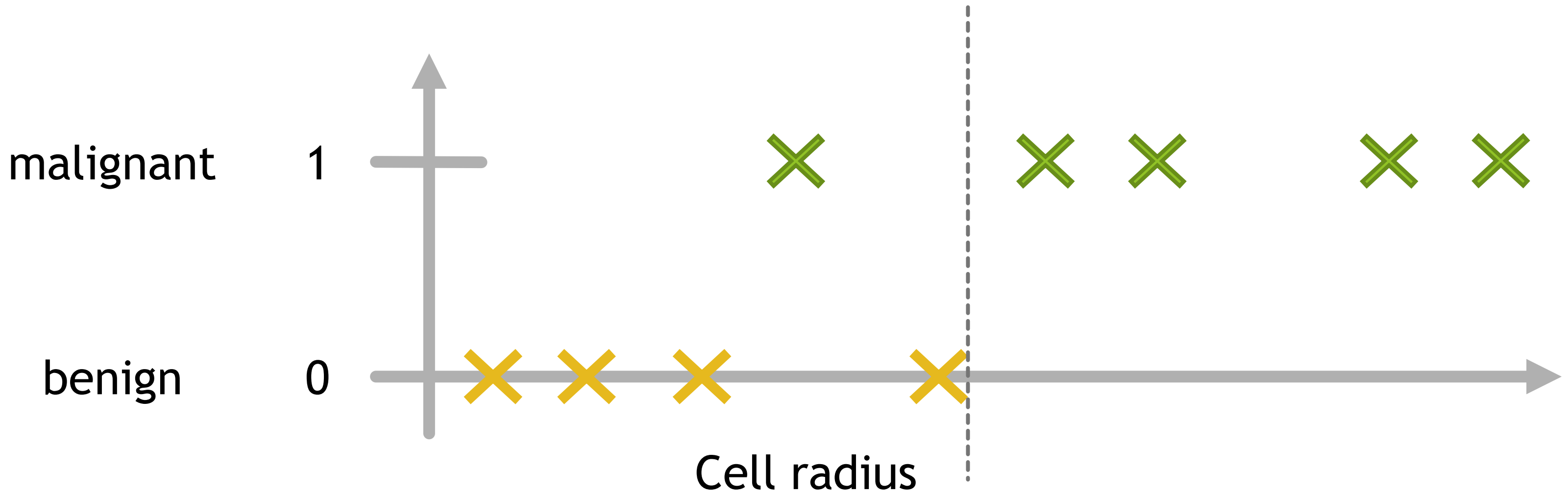
AI in Healthcare

Logistic regression

Logistic Regression Model



Logistic Regression Model

We want $0 \leq h_\theta(x) \leq 1$

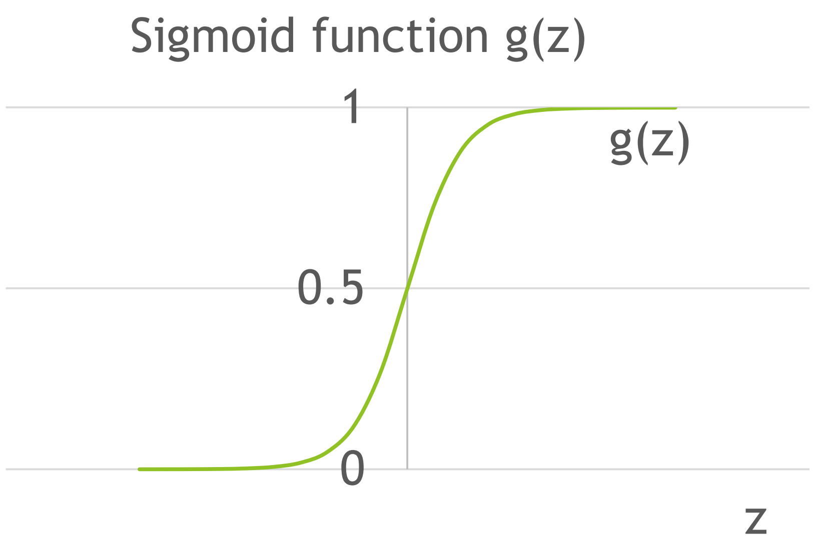
$h_\theta(x) = g(\theta^T x)$

Sigmoid function or logistic function

$g(z) = \frac{1}{1+e^{-z}}$

Therefore

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

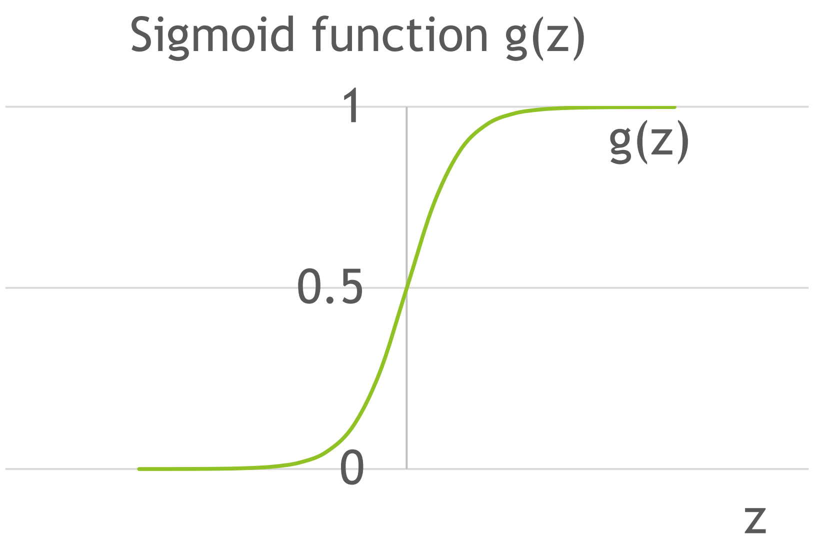


In case of one feature:

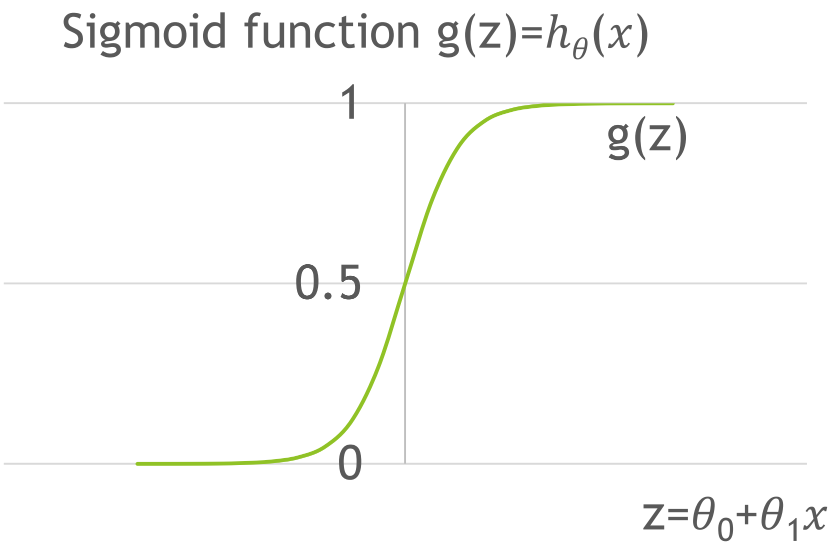
$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}$$

Hypothesis output

- The output of hypothesis $h_\theta(x)$ gives the probability that y=1
- $P(y=0|x;\theta) + P(y=1|x;\theta) = 1$
- For example:
  - predict „y=1“ if $h_\theta(x) \geq 0.5$
  - predict „y=0“ if $h_\theta(x) < 0.5$



Exercise



In case of one feature, predict „y=1“ if

- $\theta_0 + \theta_1 x < 0$
- $\theta_0 + \theta_1 x \geq 0$
- $\theta_0 + \theta_1 x < 0.5$
- $\theta_0 + \theta_1 x \geq 0.5$

Decision Boundary

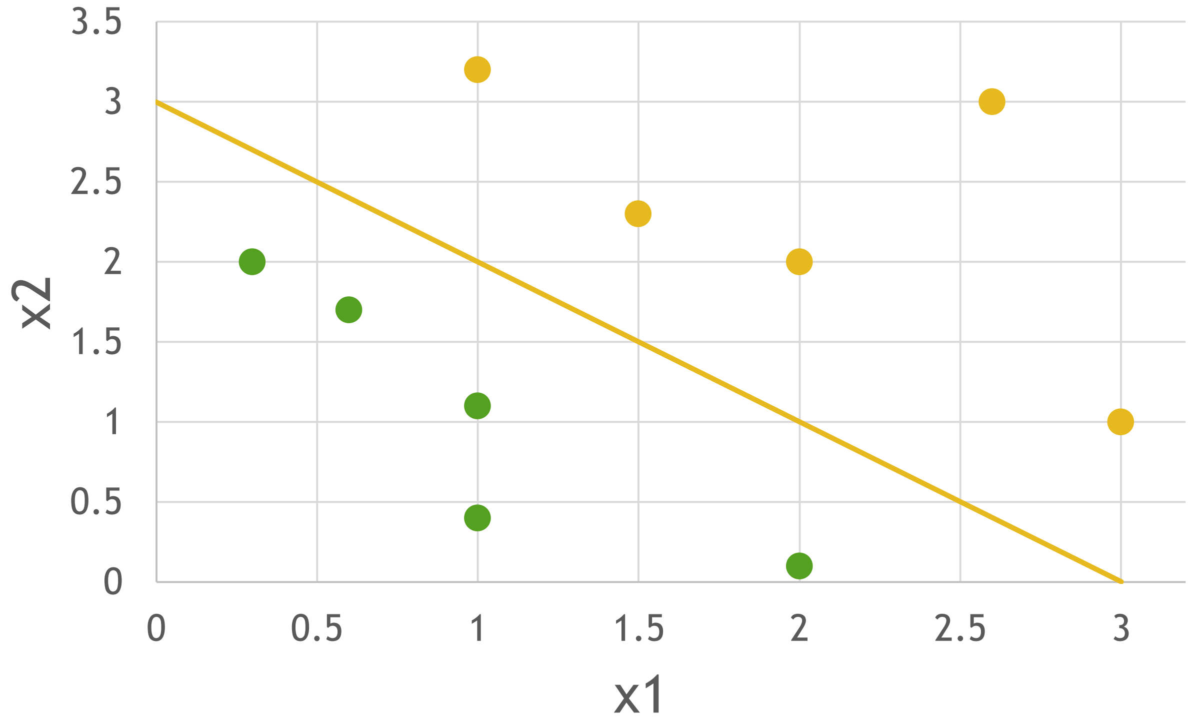
$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

Predict „y=1“ if
$\theta_0 + \theta_1 x_1 + \theta_2 x_2 \geq 0$

For example if
$\theta_0 = -3$; $\theta_1 = 1$ and $\theta_2 = 1$, we get:
Predict „y=1“ if $x_1 + x_2 \geq 3$

Therefore, we need to plot the line

$x_1 + x_2 = 3$



Exercise

- Consider logistic regression with two features x_1 and x_2. Suppose $\theta_0 = 5$, $\theta_1 = -1$ and $\theta_2 = 0$. Draw the decision boundary and show in which area „y=1“ and where „y=0“.

Logistic Regression Cost Function

- In linear regression:

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Let's say

$$cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = \frac{1}{2}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

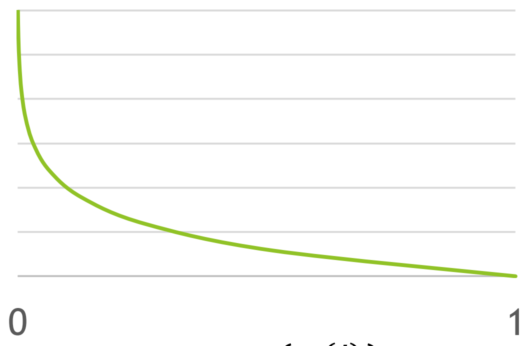
- Therefore in linear regression:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

- In logistic regression:

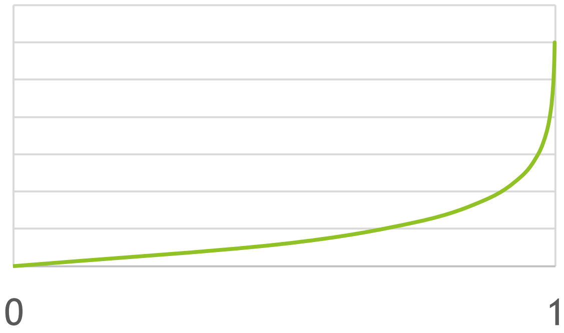
$$cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = \begin{cases} -log\left(h_\theta\left(x^{(i)}\right)\right) \; if \; y = 1 \\ -log\left(1 - h_\theta\left(x^{(i)}\right)\right) \; if \; y = 0 \end{cases}$$

$-log\left(h_\theta\left(x^{(i)}\right)\right)$ If y=1



$h_\theta\left(x^{(i)}\right)$

If y=0 $-log\left(1 - h_\theta\left(x^{(i)}\right)\right)$



$h_\theta\left(x^{(i)}\right)$

Logistic Regression Cost Function

- We have

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

$$cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = \begin{cases} -log\left(h_\theta\left(x^{(i)}\right)\right) \; if \; y = 1 \\ -log\left(1 - h_\theta\left(x^{(i)}\right)\right) \; if \; y = 0 \end{cases}$$

- We can write out cost differently:

$$cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = -y^{(i)} \cdot log\left(h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) \cdot log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

because y is always 0 or 1

- Therefore:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \cdot log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \cdot log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right]$$

Logistic Regression Gradient Descent

- Gradient descent for logistic regression:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x_j^{(i)}$$

$$j = 0, \ldots, n$$

- The same as for linear regression!
- However, in linear regression $h_\theta(x) = \theta^T x$ and in logistic regression $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Regularized Logistic Regression

- Logistic regression cost function with regularization:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \cdot log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \cdot log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

- Logistic regression gradient descent with regularization:

Repeat until convergence:

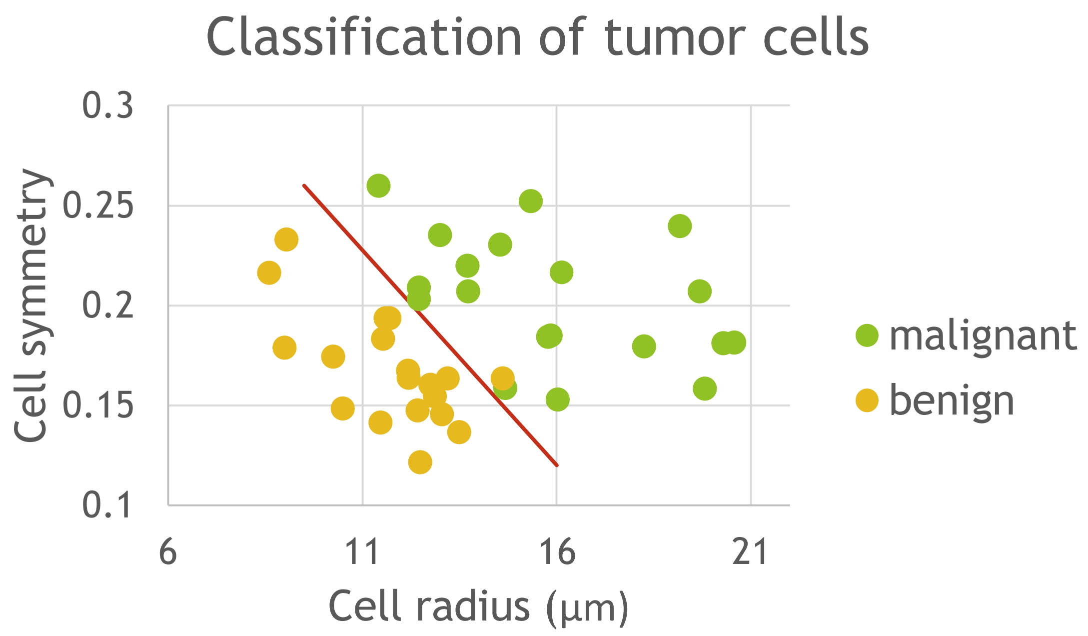
$$\theta_0 := \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha\frac{1}{m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)} + \lambda\sum_{i=1}^{n}\theta_j\right]$$

$$j = 1, \ldots, n$$

$$\text{Classification accuracy} = \frac{\text{number of correct predictions}}{\text{number of all predictions}}$$

$$\text{Misclassification error} = \frac{\text{number of mislabeled predictions}}{\text{number of all predictions}}$$



Evaluating a Hypothesis

Infant ID	Gestational Age (weeks)	Birth Weight (grams)
1	34,7	1895
2	36	2030
3	29,3	1440
4	40,1	2835
5	35.7	3090
6	42.4	3827
7	40.3	3260
8	37.3	2690
9	40.9	3285
10	38.3	2920

(1) Divide dataset:

- 70 % training set and 30 % test set or
- 60 % training set, 20 % cross validation set and 20 % test set

(2) Learn parameter θ from training set

(3) Evaluate hypothesis on test set

- Compute test set error

$$J_{test}(\theta) = -\frac{1}{m_{test}}\left[\sum_{i=1}^{m_{test}} y_{test}^{(i)} \cdot log\left(h_\theta\left(x_{test}^{(i)}\right)\right) + (1 - y_{test}^{(i)}) \cdot log\left(1 - h_\theta\left(x_{test}^{(i)}\right)\right)\right]$$

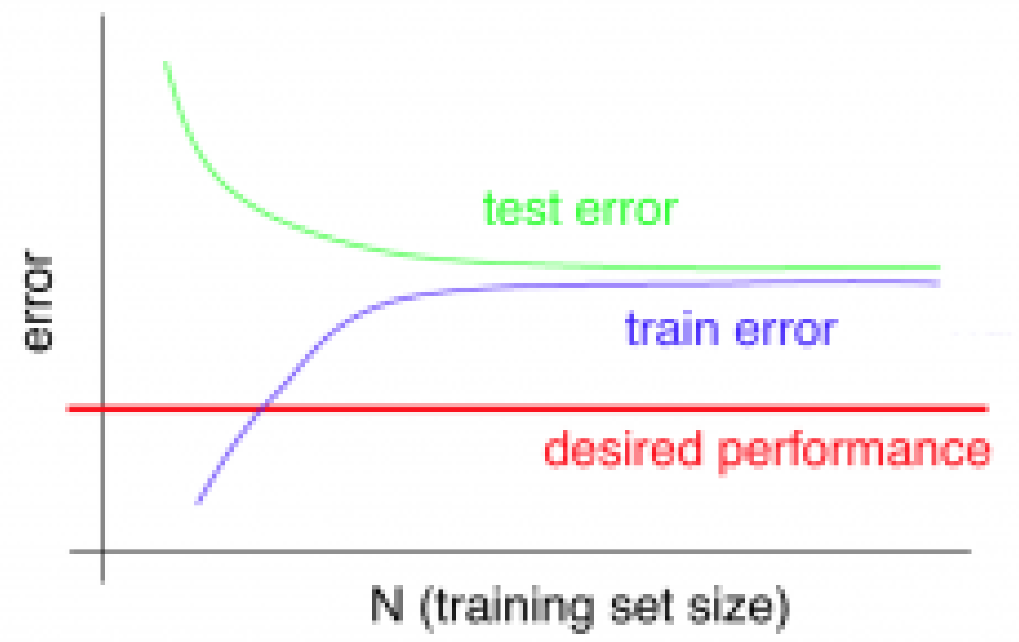
- Misclassification error
- Classification accuracy

Exercise

- If we get low training error $J(\theta)$ and high test error $J_{test}(\theta)$, what can we say about our learning algorithm hypothesis?

Evaluating a Hypothesis

Underfitting – both, test error and CV error are high



Overfitting – train error is low, but CV error is high

Images from Andrew Ng Coursera course „Machine Learning“

Evaluating a Hypothesis

- Getting more training examples - fixes overfitting
- Trying smaller sets of features - fixes overfitting
- Adding features – fixes underfitting
- Trying polynomial features - fixes underfitting
- Increasing λ – fixes overfitting
- Decreasing λ – fixes underfitting

MATLAB Assignment

You measure the cell radius 13.5 µm and cell symmetry 0.193. What is the probability of this cell being malignant?

- Create logistic regression model in MATLAB using function *mnrfit*
- Use the obtained model to estimate the probability using function *mnrval*
- In X, first row is cell radius and second row is cell symmetry. Y is 1 if tumor is malignant and 0 if benign.
- Plot data and predicted line separating malignant and benign data.