Classification
ooo

Perceptron
ooo

Logistic Regression
ooo

3 (a, b, c)
ooooo

Classification I

COMP9417, 23T1

Classification
ooo

Perceptron
ooo

Logistic Regression
ooo

3 (a, b, c)
ooooo

1. Classification

2. Perceptron

3. Logistic Regression

4. 3 (a, b, c)

Classification
●○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○○○○

Section 1

Classification

Classification

Recall the standard form of a machine learning problem:

- We have 'input' data X and targets/outputs y
- Our data can be modelled as $y = f(X)$
- Goal is to find the best approximation for f as $\hat{f}$

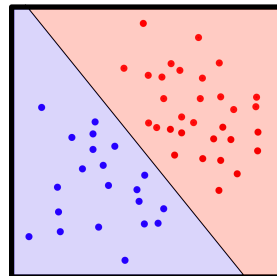Here, $f(x)$ outputs *classes* rather than numeric values.

Note:
We call a two-class problem a binary classification problem.

Linearly Separable Datasets

We define a **linearly separable** dataset as one which can be classified in a binary fashion using a hyperplane.

More simply, if you can classify it by drawing a line through it. Your dataset is linearly separable.

Classification
○○○

Perceptron
●○○

Logistic Regression
○○○

3 (a, b, c)
○○○○○

Section 2

Perceptron

Classification
ooo

Perceptron
o●o

Logistic Regression
ooo

3 (a, b, c)
ooooo

Perceptron

Learns weights w for a decision boundary $w^T \mathbf{x} = 0$, where $\mathbf{x}$ represents points on the Cartesian plane, not our dataset.

Key Properties

- The classic perceptron solves only binary classification
- **Always converges** to a solution if the dataset in linearly separable
- Solutions can differ depending on starting weights and learning rate

The algorithm

For weights w and a learning rate η.

$\quad converged \leftarrow 0$

$\quad$ **while** not $converged$ **do**

$\quad\quad converged \leftarrow 1$

$\quad\quad$ **for** $x_i \in X, y_i \in y$ **do**

$\quad\quad\quad$ **if** $y_i w \cdot x_i \leq 0$ **then**

$\quad\quad\quad\quad w \leftarrow w + \eta y_i x_i$

$\quad\quad\quad\quad converged \leftarrow 0$

$\quad\quad\quad$ **end if**

$\quad\quad$ **end for**

$\quad$ **end while**

Classification
000

Perceptron
000

Logistic Regression
●○○

3 (a, b, c)
00000

Section 3

Logistic Regression

Logistic Regression

Often called *logit* **model**. A way for us to use a linear combination $w^T x$ to predict probabilities of a binary classification problem.

For a data point (x_i, y_i) the model will predict:
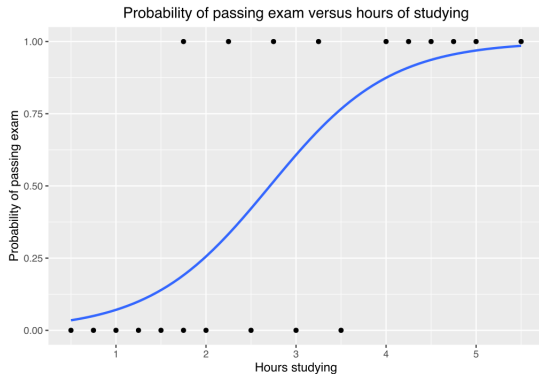
$$P(y_i = 1 | x_i)$$

Simply, the probability that the target belongs to class 1 given the datapoint at index i.

Classification
○○○

Perceptron
○○○

Logistic Regression
○○●

3 (a, b, c)
○○○○○

The logistic regression is defined as the following function:

$$\sigma(w^T x_i) = \frac{1}{1 + e^{-w^T x_i}}$$

In the basic case where we only have one feature:

$$\sigma(w^T x_i) = \frac{1}{1 + e^{-w_0 - w_1 x_i}}$$

Probability of passing exam versus hours of studying

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
●○○○○

Section 4

3 (a, b, c)

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○●○○○○

3 (a, b, c)

If we define the binary prediction problem as a probability:

$$P(y = 1|x) = p(x)$$

We write the logistic regression prediction as:

$$\hat{p}(x) = \sigma(\hat{w}^T x)$$

$$\text{where } \sigma(z) = \frac{1}{1 + e^{-z}}$$

where we predict the class of an input x to be 1 if $\hat{p}(x) \geq 0.5$.

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○●○○

3a

What is the role of the sigmoid function here?

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○●○○

3a

What is the role of the sigmoid function here?

In a linear model, we can't simply predict probabilities or classes with the classic equation $\hat{p}(x) = \hat{w}^T x$.

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○●○○

3a

What is the role of the sigmoid function here?

In a linear model, we can't simply predict probabilities or classes with the classic equation $\hat{p}(x) = \hat{w}^T x$. The sigmoid $\sigma(z)$ us model probabilities in a valid interval ($[0, 1]$).

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○○●○

3b

Consider the statistical view of the binary classification problem $y_i|x_i \sim \text{Bernoulli}(p_i^*)$ where $p_i^* = \sigma(x_i^T w)$ is our logistic regression model.

By definition of the Bernoulli:

$$P(y|x) = p^y(1-p)^{1-y}$$

So, we can estimate p using MLE:

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○○●○

3b

Consider the statistical view of the binary classification problem $y_i | x_i \sim \text{Bernoulli}(p_i^*)$ where $p_i^* = \sigma(x_i^T w)$ is our logistic regression model.

By definition of the Bernoulli:

$$P(y|x) = p^y (1-p)^{1-y}$$

So, we can estimate p using MLE:

$$\ln L(w) = \ln \left(\prod_{i=1}^{n} P(y_i|x_i) \right)$$

Classification
ooo

Perceptron
ooo

Logistic Regression
ooo

3 (a, b, c)
oooeo

3b

Consider the statistical view of the binary classification problem $y_i|x_i \sim \text{Bernoulli}(p_i^*)$ where $p_i^* = \sigma(x_i^T w)$ is our logistic regression model.

By definition of the Bernoulli:

$$P(y|x) = p^y(1-p)^{1-y}$$

So, we can estimate p using MLE:

$$\ln L(w) = \ln \left(\prod_{i=1}^{n} P(y_i|x_i) \right)$$

$$= \sum_{i=1}^{n} \ln P(y_i|x_i)$$

Classification
000

Perceptron
000

Logistic Regression
000

3 (a, b, c)
0000●

$$= \sum_{i=1}^{n} \ln \left(p_i^{y_i} (1 - p_i)^{1-y_i} \right)$$

$$= \sum_{i=1}^{n} \ln \left(p_i^{y_i} (1 - p_i)^{1-y_i} \right)$$

$$= \sum_{i=1}^{n} \left[y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \right]$$

$$= \sum_{i=1}^{n} \ln \left(p_i^{y_i} (1 - p_i)^{1-y_i} \right)$$

$$= \sum_{i=1}^{n} \left[y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln \left(\sigma(w^T x_i) \right) + (1 - y_i) \ln \left(1 - \sigma(w^T x_i) \right) \right]$$

Classification
○○○

Perceptron
○○○

Logistic Regression
○○○

3 (a, b, c)
○○○○●

$$= \sum_{i=1}^{n} \ln \left(p_i^{y_i} (1 - p_i)^{1-y_i} \right)$$

$$= \sum_{i=1}^{n} \left[y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln \left(\sigma(w^T x_i) \right) + (1 - y_i) \ln \left(1 - \sigma(w^T x_i) \right) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln \left(\frac{\sigma(w^T x_i)}{1 - \sigma(w^T x_i)} \right) + \ln \left(1 - \sigma(w^T x_i) \right) \right]$$