

# Machine Learning

## Lecture 3 - Linear Modelling...again

Ke Yuan  
Ke.Yuan@glasgow.ac.uk

School of Computing Science  
University of Glasgow

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

## Multiplication

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

## Multiplication

$$\begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$\mathbf{w}^T \mathbf{x} = [w_0, w_1]$$

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

## Multiplication

$$\mathbf{w}^T \mathbf{x} = [w_0 \ w_1] \begin{bmatrix} 1 \\ x \end{bmatrix}$$

Diagram illustrating the multiplication: A red bracket connects the vector  $[w_0 \ w_1]$  to the matrix  $\begin{bmatrix} 1 \\ x \end{bmatrix}$ . Red arrows point from the elements  $w_0$  and  $w_1$  to the corresponding columns of the matrix.

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

## Multiplication

$$\mathbf{w}^T \mathbf{x} = [w_0, w_1] \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$w_0$   $w_1$   $x$

$(w_0 \times 1) + (w_1 \times x)$

# Our model in vector form

## Aim

To write our model  $t = w_0 + w_1x$  in terms of  $\mathbf{w}$  and  $\mathbf{x}$ .

We'll need some ways of manipulating vectors:

## Transpose

Rotates the vector:

$$\mathbf{w}^T = [w_0, w_1], \mathbf{x}^T = [1, x]$$

## Multiplication

$$\mathbf{w}^T \mathbf{x} = [w_0, w_1] \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$(w_0 \times 1) + (w_1 \times x)$

$$\mathbf{w}^T \mathbf{x} = w_0 + w_1x$$
$$\mathbf{a}^T \mathbf{b} = \sum_{d=1}^D a_d b_d$$

# Encapsulate data in a vector

- ▶ We can make a data vector too:

# Encapsulate data in a vector

- ▶ We can make a data vector too:

$$x^1$$

# Encapsulate data in a vector

- ▶ We can make a data vector too:

 $x^0$  $x^1$

# Encapsulate data in a vector

- ▶ We can make a data vector too:

$$\mathbf{x} = \begin{bmatrix} x^0 \\ x^1 \end{bmatrix}$$

# Matrix multiplication

$$\mathbf{X}\mathbf{w} =$$

$$\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix}$$

# Matrix multiplication

$\mathbf{Xw} =$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix} (1 \times w_0) \quad \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

The diagram illustrates the multiplication of a vector  $x$  by a scalar  $w_0$ . The vector  $x$  is shown as a column vector with entries 1,  $x_1$ ,  $\vdots$ , and  $X_N$ . The scalar  $w_0$  is shown as a row vector with entries  $w_0$  and  $w_1$ . Red arrows indicate the scaling of each component of  $x$  by  $w_0$ : one arrow points from the first entry '1' in  $x$  to the first entry  $w_0$  in the scalar, and another arrow points from the second entry  $x_1$  in  $x$  to the second entry  $w_1$  in the scalar.

# Matrix multiplication

$\mathbf{Xw} =$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & X_N \end{bmatrix} (1 \times w_0) + (x_1 \times w_1)$$

The diagram shows a red arrow originating from the circled  $x_1$  in the input vector and pointing to the circled  $w_1$  in the weight vector, indicating the calculation of the first term in the sum.

# Matrix multiplication

$\mathbf{Xw} =$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots \\ 1 & X_N \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$(1 \times w_0) + (x_1 \times w_1)$

$(1 \times w_0) + (x_2 \times w_1)$

# Matrix multiplication

$$\mathbf{X}\mathbf{w} =$$

$$\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix} (1 \times w_0) + (x_1 \times w_1)$$
$$(1 \times w_0) + (x_2 \times w_1)$$
$$(1 \times w_0) + (x_N \times w_1)$$

$$= \begin{bmatrix} \mathbf{w}^T \mathbf{x}_1 \\ \mathbf{w}^T \mathbf{x}_2 \\ \vdots \\ \mathbf{w}^T \mathbf{x}_N \end{bmatrix}$$

# Partial diff. wrt vector

The result of taking the partial derivative with respect to a vector is a vector where each element is the partial derivative with respect to one parameter:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w_0} \\ \frac{\partial \mathcal{L}}{\partial w_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial w_K} \end{bmatrix}$$

# Partial diff. wrt vector

The result of taking the partial derivative with respect to a vector is a vector where each element is the partial derivative with respect to one parameter:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w_0} \\ \frac{\partial \mathcal{L}}{\partial w_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial w_K} \end{bmatrix}$$

## Useful identities:

$f(\mathbf{w})$	$\frac{\partial f}{\partial \mathbf{w}}$
$\mathbf{w}^\top \mathbf{x}$	$\mathbf{x}$
$\mathbf{x}^\top \mathbf{w}$	$\mathbf{x}$
$\mathbf{w}^\top \mathbf{w}$	$2\mathbf{w}$
$\mathbf{w}^\top \mathbf{Cw}$	$2\mathbf{Cw}$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}} \left( \frac{1}{N} (\mathbf{t} - \mathbf{Xw})^\top (\mathbf{t} - \mathbf{Xw}) \right) &= \frac{1}{N} (2\mathbf{X}^\top \mathbf{Xw} - 2\mathbf{X}^\top \mathbf{t}) \\ \mathbf{X}^\top \mathbf{Xw} &= \mathbf{X}^\top \mathbf{t}\end{aligned}$$

## Matrix transpose

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}, \quad \mathbf{X}^\top = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix}$$

## Transpose of sum/product

$$(\mathbf{a} + \mathbf{b})^\top = \mathbf{a}^\top + \mathbf{b}^\top, \quad (\mathbf{Xw})^\top = \mathbf{w}^\top \mathbf{X}^\top$$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}} \left( \frac{1}{N} (\mathbf{t} - \mathbf{Xw})^\top (\mathbf{t} - \mathbf{Xw}) \right) &= \frac{1}{N} (2\mathbf{X}^\top \mathbf{Xw} - 2\mathbf{X}^\top \mathbf{t}) = \mathbf{0} \\ \mathbf{X}^\top \mathbf{Xw} &= \mathbf{X}^\top \mathbf{t}\end{aligned}$$

## Matrix transpose

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}, \quad \mathbf{X}^\top = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix}$$

## Transpose of sum/product

$$(\mathbf{a} + \mathbf{b})^\top = \mathbf{a}^\top + \mathbf{b}^\top, \quad (\mathbf{Xw})^\top = \mathbf{w}^\top \mathbf{X}^\top$$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\mathbf{X}^T \mathbf{X} \mathbf{w} = \mathbf{X}^T \mathbf{t}$$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\mathbf{X}^T \mathbf{X} \mathbf{w} = \mathbf{X}^T \mathbf{t}$$

## Matrix inverse

Inverse is defined (for a square matrix  $\mathbf{A}$ ) as the matrix  $\mathbf{A}^{-1}$  that satisfies:

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$$

Where  $\mathbf{I}$  is the *identity* matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \text{ and } \mathbf{I}\mathbf{A} = \mathbf{A}, \text{ for any } \mathbf{A}$$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\begin{aligned}\mathbf{X}^T \mathbf{X} \mathbf{w} &= \mathbf{X}^T \mathbf{t} \\ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \mathbf{w} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}\end{aligned}$$

## Matrix inverse

Inverse is defined (for a square matrix  $\mathbf{A}$ ) as the matrix  $\mathbf{A}^{-1}$  that satisfies:

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$$

Where  $\mathbf{I}$  is the *identity* matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \text{ and } \mathbf{I} \mathbf{A} = \mathbf{A}, \text{ for any } \mathbf{A}$$

# Computing $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

$$\begin{aligned}\mathbf{X}^T \mathbf{X} \mathbf{w} &= \mathbf{X}^T \mathbf{t} \\ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \mathbf{w} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t} \\ \mathbf{w} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}\end{aligned}$$

## Matrix inverse

Inverse is defined (for a square matrix  $\mathbf{A}$ ) as the matrix  $\mathbf{A}^{-1}$  that satisfies:

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$$

Where  $\mathbf{I}$  is the *identity* matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \text{ and } \mathbf{I} \mathbf{A} = \mathbf{A}, \text{ for any } \mathbf{A}$$