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 $t = w_0 + w_1 x$ 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 $t = w_0 + w_1 x$ in terms of $\mathbf{w}$ and $\mathbf{x}$.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Our model in vector form

Aim

To write our model $t = w_0 + w_1 x$ in terms of $\mathbf{w}$ and $\mathbf{x}$.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Multiplication

Our model in vector form

Aim

To write our model $t = w_0 + w_1 x$ in terms of **w** and **x**.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Multiplication

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

Our model in vector form

Aim

To write our model $t = w_0 + w_1 x$ in terms of $\mathbf{w}$ and $\mathbf{x}$.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Multiplication

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

Our model in vector form

Aim

To write our model $t = w_0 + w_1 x$ in terms of $\mathbf{w}$ and $\mathbf{x}$.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Multiplication

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

Our model in vector form

Aim

To write our model $t = w_0 + w_1 x$ in terms of $\mathbf{w}$ and $\mathbf{x}$.

We'll need some ways of manipulating vectors:

Transpose

Rotates the vector:

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

Multiplication

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

$$\mathbf{w}^\mathsf{T}\mathbf{x} = w_0 + w_1 x$$

$$\mathbf{a}^\mathsf{T}\mathbf{b} = \sum_{d=1}^{D} a_d b_d$$

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:

$$\begin{matrix} x^0 \\ x^1 \end{matrix}$$

Encapsulate data in a vector

- We can make a data vector too:

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

Matrix multiplication

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

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

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

Matrix multiplication

$\mathbf{Xw} =$

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

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

Matrix multiplication

$$\mathbf{Xw} =$$

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

Matrix multiplication

$\mathbf{Xw} =$

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

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

Matrix multiplication

$\mathbf{Xw} =$

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

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

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

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 identites:

$f(\mathbf{w})$	$\frac{\partial f}{\partial \mathbf{w}}$
$\mathbf{w}^T\mathbf{x}$	$\mathbf{x}$
$\mathbf{x}^T\mathbf{w}$	$\mathbf{x}$
$\mathbf{w}^T\mathbf{w}$	$2\mathbf{w}$
$\mathbf{w}^T\mathbf{C}\mathbf{w}$	$2\mathbf{C}\mathbf{w}$

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

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

Matrix transpose

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

Transpose of sum/product

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

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

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

Matrix transpose

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

Transpose of sum/product

$$(\mathbf{a} + \mathbf{b})^{\mathsf{T}} = \mathbf{a}^{\mathsf{T}} + \mathbf{b}^{\mathsf{T}}, \ (\mathbf{X}\mathbf{w})^{\mathsf{T}} = \mathbf{w}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}$$

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

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

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

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

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

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

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

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

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,

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