Lecture theatre changes

Tomorrow: Appleton Tower

(Thu 19)

LT 5

Also Thursday 26th

Wednesdays:

(Every week after today)

David Hume Tower LTs

LTA

Shane Legg
@ShaneLegg

Learn:
1. linear algebra well (e.g. matrix math)
2. calculus to an ok level (not advanced stuff)
3. prob. theory and stats to a good level
4. theoretical computer science basics
5. to code well in Python and ok in C++

Then read and implement ML papers and *play* with stuff! :-)

> **aron** @aron65900682
> @ShaneLegg Hey Shane I'm currently 17 from London England and am very passionate about AI, also learning about in-depth human needs. What would be the 5 pieces of advice and tips you would give to a young person like me?

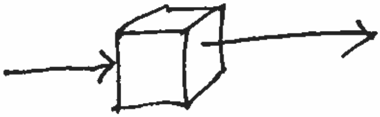
2:44 PM - 31 Jul 2018

Cartoon view of Machine Learning

Input (pixels)

$\underline{x}$

$\mathbf{x}$ or $\vec{x}$



function, f

Location of face

Observed output

$\underline{y} = [a, b, w, h]^T$

Email / text $\longrightarrow$ Label $\quad f(x) \in \{spam, ok, phishing\}$

$\underline{x}$

Write f by hand?

If "Ray-Ban" in $\underline{x}$: spam += 10

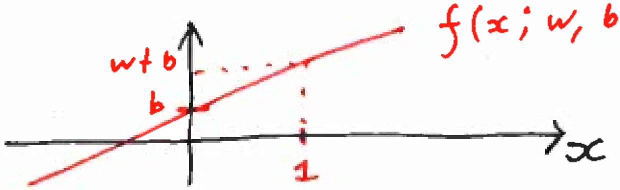
If "IT Help Desk" in $\underline{x}$: phishing += 100

If "Bayesian" in $\underline{x}$: ok += 10^6

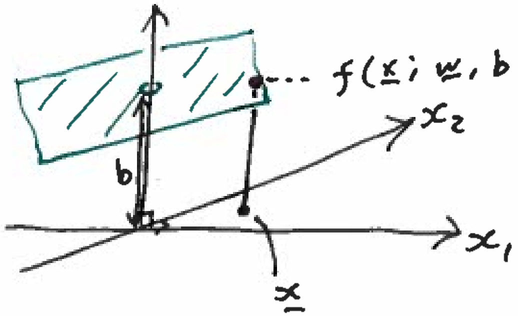
Parameters $\underline{\theta}$ or $\underline{w}$ weights

Prediction = argmax(spam, phishing, ok)

Linear Functions



$$f(x; w, b) = wx + b$$

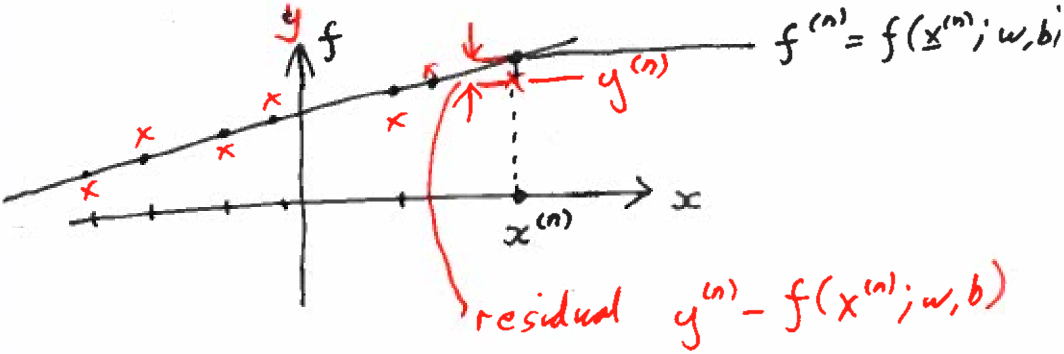


$$f(\underline{x}; \underline{w}, b) = w_1 x_1 + w_2 x_2 + b$$

$$= \underline{w}^T \underline{x} = \underline{x}^T \underline{w}$$

Data

$$\{(\underline{x}^{(n)}, y^{(n)})\}_{n=1}^{N}$$



In vector and matrix format:

$$\underline{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} \qquad X = \begin{bmatrix} \text{---} & \underline{x}^{(1)\top} & \text{---} \\ \text{---} & \underline{x}^{(2)\top} & \text{---} \\ & \vdots & \\ x_1^{(N)} & x_2^{(N)} \; \cdots & x_D^{(N)} \end{bmatrix}$$

D-dimensional regression

Matlab: $N \times 1$ matrix

Python $(N,)$ array (vector)

$(N, 1)$ array

If y.shape $= (N,)$

y[:, None] is $(N, 1)$ array

$$\underline{f} = \begin{bmatrix} f(\underline{x}^{(1)}; \underline{w}, b) \\ \vdots \\ f(\underline{x}^{(N)}; \underline{w}, b) \end{bmatrix}$$

Least squares fitting

Minimize $\displaystyle\sum_{n=1}^{N} \left(y^{(n)} - f(\underline{x}^{(n)}; \underline{w}, b)\right)^2$

Minimize $(\underline{y} - \underline{f})^T (\underline{y} - \underline{f})$

Model with zero intercept $(b=0)$

$f(\underline{x}; \underline{w}) = \underline{w}^T \underline{x} = \underline{x}^T \underline{w}$

Linear map

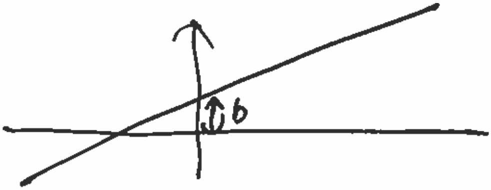
$g(\underline{x} + \underline{z}) = g(\underline{x}) + g(\underline{z})$

$g(c\,\underline{x}) = c\,g(\underline{x})$

$$\underset{N\times 1}{\underline{f}} = \underset{N\times D}{X}\;\underset{D\times 1}{\underline{w}} \approx \underline{y}$$

Matlab: `w_fit = X\y;`

Python: `= np.linalg.lstsq(X, y)[0]`



$$\tilde{X} = \begin{bmatrix} — \mathbf{x}^{(1)\top} — & 1 \\ — \mathbf{x}^{(2)\top} — & 1 \\ \vdots & \\ — \mathbf{x}^{(N)\top} — & 1 \end{bmatrix} = \overbrace{\begin{bmatrix} X & \begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix} \end{bmatrix}}^{D+1 \text{ wide}}$$

$$\tilde{\mathbf{w}} = \arg\min \| \mathbf{y} - \tilde{X} \underbrace{\tilde{\mathbf{w}}}_{(D+1)\text{-dim vector}} \|^2$$

Fit $\mathbf{y}$ with $\mathbf{f} = \tilde{X}\tilde{\mathbf{w}} = X \underbrace{\tilde{\mathbf{w}}_{1:D}}_{\mathbf{w}} + \underbrace{\tilde{w}_{D+1}}_{b}$

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

Please tell me about mistakes during the lecture

$$\underline{f} = \Phi \, \underline{w}$$

Φ: $N \times K$

$\underline{w}$: $K \times 1$ vector of parameters

Each row still a datapoint

$$\Phi = \begin{bmatrix} \text{---} & \underline{\phi}(\underline{x}^{(1)})^T & \text{---} \\ & \vdots & \\ \text{---} & \underline{\phi}(\underline{x}^{(N)})^T & \text{---} \end{bmatrix}$$

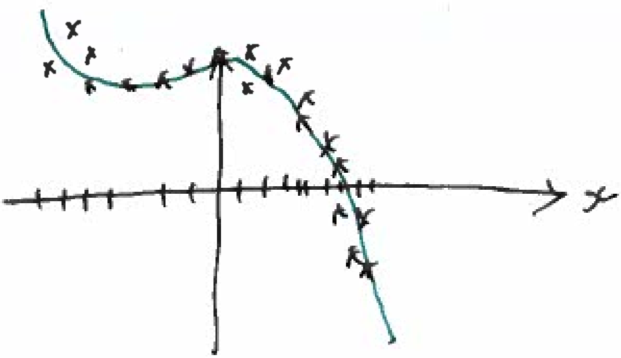
<u>Example</u> (1D example)

$$\underline{\phi}(\underline{x}) = [\,1 \quad x \quad x^2 \quad x^3\,]^T$$

Fit $\underline{y} \approx \underline{f} = \Phi \underline{w}$ (by "linear regression")

$$f(\underline{x}) = \underline{w}^T \underline{\phi}(\underline{x})$$

$$= w_1 + w_2 x + w_3 x^2 + w_4 x^3$$



What about $D > 1$?

Vector inputs $\underline{x}$

$$\underline{\Phi} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \left(x_1^{(1)}\right)^2 & x_1^{(1)} x_3^{(1)} \\ & & & & & \\ & & & & & \end{bmatrix}$$