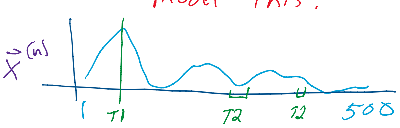
315 Lecture 4

- Warmup: Simplify or expand these expressions
  - $e^{z_1} e^{z_2}$ $\quad e^{z_1 + z_2}$
  - $\log e^z$ $\quad z$
  - $\log(z_1 z_2)$ $\quad \log z_1 + \log z_2$
  - $\log\left(\frac{z_1}{z_2}\right)$ $\quad \log z_1 - \log z_2$
  - $\frac{d}{dz}\frac{1}{z}$ $\quad -\frac{1}{z^2}$
  - $\frac{d}{dz}\log(z)$ $\quad \frac{1}{z}$

- Announcements
  - HW2
    - Due 2/4, 11:59 pm
    - Written + Programming
    - Submit both to Gradescope
- Plan
  - Today
    - MRI example
    - Neural network primer
    - MLE
    - MLE for linear regression
  - Wed
    - Logistic regression

$$\mathcal{D} = \{\vec{x}^{(n)}, \vec{y}^{(n)}\}_{n=1}^{N} \qquad \vec{x} \in \mathbb{R}^{500\times 1} \qquad \vec{y} \in \mathbb{R}^2$$

Poll 1: Could a linear model accurately model this?



$\rightarrow (T1, T2)$

$W x \rightarrow y \begin{bmatrix} T1 \\ T2 \end{bmatrix}$

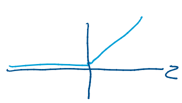
$$W \begin{bmatrix} 0\,0\,0\,1\,0\,0\,0\,0 & \cdots & 0 \\ 0\,0\,0 & \cdots & 0\,1\,1\,0\,0\,0\,1\,0\,0\,0\,0 \end{bmatrix} \begin{array}{l} \rightarrow T1 \\ \rightarrow T2 \end{array}$$

Neural Networks

$\text{Input}_{500} \rightarrow \text{Linear}_2 \rightarrow \overset{\hat{y}}{\begin{bmatrix} T1 \\ T2 \end{bmatrix}}$ } MSE, SGD

$\text{Input}_{500} \rightarrow \text{Linear}_{250} \xrightarrow[250]{} \text{Linear}_{125} \rightarrow \text{Linear}_2$ } MSE SGD

$\text{ReLU}(z) = \max(z, 0)$



$\text{ReLU}(\vec{z})$

$\text{Input}_{500} \rightarrow \text{Linear}_{250} \rightarrow \text{ReLU} \rightarrow \text{Linear}_{125} \rightarrow \text{ReLU}$
$\rightarrow \text{Linear}_2 \rightarrow \hat{y}$ } MSE, SGD

MR Fingerprinting Assumptions

MR Fingerprinting Assumptions

- $x^{(n)}, y^{(n)}$ iid   i.e. pixels are iid
- $\epsilon \sim \mathcal{N}(\mu, \alpha^2)$

$$f(\epsilon) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{\left(-\frac{(\epsilon-\mu)^2}{2\alpha^2}\right)}$$

- Training data represents all future patients!!

Poll 2 What is the "true" meaning of Probability

Ex Coin flips

A, The number of Heads in 1000 flips

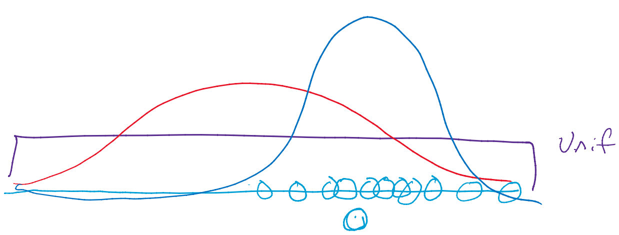
B. Uncertainty about the next flip

[Either/Both. Philisophical difference between Frequntist (A) and Bayesian (B) thinking.]

A. $P(D \mid \theta)$

B. $\underbrace{P(\theta \mid D)}_{\text{posterior}} \propto \underbrace{P(D \mid \theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{Prior}}$

Density Estimation



$$f(x=65 \mid Unif(0, 100)) = \frac{1}{100}$$

$$f(x=65 \mid \mathcal{N}(50, 25))$$

$$f(x^{(n)} \mid \mu, \alpha^2) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{\left(-\frac{(x^{(n)}-\mu)^2}{2\alpha^2}\right)}$$

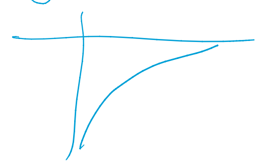
$$f(x^{(n)}|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\left(-\frac{(x^{(n)}-\mu)}{2\sigma^2}\right)}$$

$$f(\vec{x}\,|\mu,\sigma^2) = f(x^{(1)}|\mu^{\cancel{(1)}},\sigma^{2\cancel{(1)}}) \cdot f(x^{(2)}|\mu^{\cancel{(2)}},\sigma^{2\cancel{(2)}}) \cdot \ldots$$

$$= \prod_{n=1}^{N} f(x^{(n)}|\mu,\sigma^2)$$

$$= \frac{1}{2\pi\sigma^{2\,N/2}}\, e^{\left(-\frac{\sum_{n=1}^{N}(x^{(n)}-\mu)^2}{2\sigma^2}\right)}$$

log is monotonic



$z_1 < z_2$

$\log z_1 < \log z_2$

$$\hat{\theta} = \arg\max_{\theta}\, p(D|\theta)$$

$$\hat{\mu}, \hat{\sigma}^2 = \arg\max_{\mu\,\sigma^2}\, f(\vec{x}\,|\mu,\sigma^2)$$

?

$$= \arg\max\, \log f(\vec{x}\,|\mu\,\sigma^2)$$

iid

$$= \arg\max\, \log \prod_{n}^{N} f(x^{(n)}|\mu\,\sigma^2)$$

$\log z_1 z_2 = \log z_1 + \log z_2$

$$= \arg\max \sum_{n}^{N} \log f(x^{(n)}|\mu,\sigma^2)$$

$\boxed{\text{Likelihood } \; L(\theta) \;,\; L(\mu,\sigma^2)}$

$\boxed{\text{log-likelihood } \; \ell(\theta) \;,\; \ell(\mu,\sigma^2)}$

MLE of μ

$$L(\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}}\, e^{\left(-\frac{\sum_{n=1}^{N}(x^{(n)}-\mu)^2}{2\sigma^2}\right)}$$

$$\frac{\partial \ell}{\partial \theta} = 0$$

$$\frac{\partial \ell}{\partial \mu} = 0$$

$$\ell(\mu) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{\sum_{n=1}^{N}(x^{(n)}-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{\sigma^2}\sum(x^{(n)}-\mu) = 0$$

$$\sum x^{(n)} = N\hat{\mu}$$

$$\hat{\mu}_{ML} = \frac{1}{N}\sum x^{(n)}$$

MLE of σ^2

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^{2\,2}}\sum(x^{(n)}-\mu)^2$$

$$\frac{N}{\;} - \frac{1}{\;}\sum (x^{(n)}-\mu)^2$$

$\partial \sigma^2$ $\partial \sigma$ $\partial \sigma$

$$\frac{N}{2\sigma^2} = \frac{1}{2\sigma^4} \sum \left(x^{(n)} - \mu\right)^2$$

$$\hat{\sigma}^2_{ML} = \frac{1}{N} \sum \left(x^{(n)} - \hat{\mu}_{ML}\right)^2$$