PROBABILITY HW

CDF (CUMULATIVE DIST. FUN)

$$F_X(t) = \mathbb{P}[x \le t]$$

a) y_i ?

$$\begin{aligned} F_{y_i}(t) &= \mathbb{P}[y_i \le t] \\ &= \mathbb{P}[f_\Theta(x_i) + \varepsilon_i \le t] \\ &= \mathbb{P}[\varepsilon_i \le t - f_\Theta(x_i)] \\ &= F_{\varepsilon_i}(t - f_\Theta(x_i)) \end{aligned}$$

$$y_i \sim \mathcal{N}(f_\Theta(x_i), \sigma^2)$$

CUMULATIVE DIST. FUNCTION

$$F_X(t) = \mathbb{P}[X \le t]$$

PROBABILITY DENSITY FUN.

$$f_X(t) = \frac{\partial F_X(t)}{\partial t}$$

Both CDF and PDF uniquely determine the distribution! CDF is often easier to compute

Prob. of X being in $[a,b]$

$$\int_a^b f_X(t)\,dt = F_X(b) - F_X(a)$$

b) $like(\theta) = \mathbb{P}[y_1, y_2 \dots y_n | x_1 \dots x_n, \theta]$

$$= \prod_{i=1}^{n} \mathbb{P}[y_i \mid x_{i1} \dots x_{in}, \theta] =$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - f_\theta(x_i))^2}{2\sigma^2}}$$

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\text{loglike}(\Theta) = \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - f_\Theta(x_i))^2}{2\sigma^2}}\right]$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \underbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f_\Theta(x_i))^2}_{\text{Mean Squared Error}}$$

c) Maximizing loglikelihood is equivalent to minimizing loss function (for example, mean squared error).

2.

$$\text{loglike}(\Theta) =$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\log(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{n}(y_i - f_\Theta(x_i))^2/\sigma_i^2$$

c) The errors aren't identically distributed:
the variance depends on the sample.

A big deviation (epsilon) when the variance is small decreases the likelihood much more than a big deviation when the variance itself is large.

3.

theta - probability of getting H in one toss

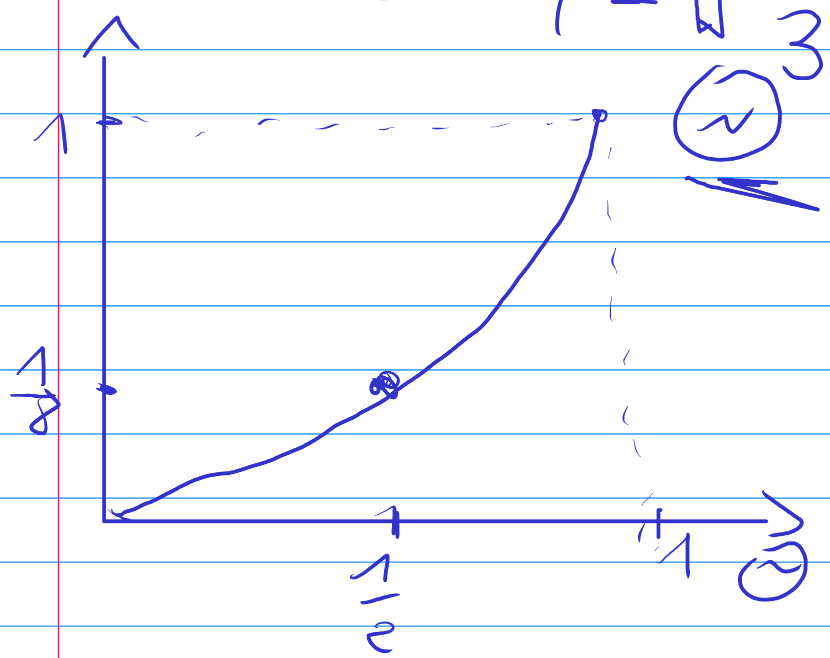
like(theta) = P[HHH | theta] = theta^3 } likelihood

MLE: $\max$ like(θ) ?

ONE WAY:

$$\frac{d}{d\theta} \text{like}(\theta) = 3\theta^2 \overset{?}{=} 0$$

$$\theta \in [0,1]$$



$$\theta^* = 1$$

Mode, mean

↳ we need to compute posterior distribution

$$\text{mode} = \arg\max P(\theta \mid HHH_c, \alpha, \beta)$$

↳ in Beta distribution

$$\text{mode} = \frac{\alpha - 1}{\alpha + \beta - 2} \qquad \text{Beta}(\alpha, \beta)$$

$$\text{mean} = \frac{\alpha}{\alpha + \beta}$$

These are only for Beta dist

Bayes rule:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\underbrace{p(\theta \mid HHH; \alpha, \beta)}_{\text{posterior}} = ?$$

$$p(\theta \mid HHH; \alpha, \beta) = \frac{p(HHH \mid \theta; \alpha, \beta)\, p(\theta \mid \alpha, \beta)}{\underbrace{P[HHH \mid \alpha, \beta]}_{\text{const}}}$$

$$= \frac{\theta^3 \cdot \mathrm{Beta}(\theta; \alpha, \beta)}{Z} = \ldots$$

$$\boxed{\mathrm{Beta}(\theta; \alpha, \beta) = \theta^{\alpha-1} (1-\theta)^{\beta-1}}$$

Beta distribution

$$= \frac{1}{Z} \theta^3 \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \frac{1}{Z} \theta^{\alpha+2} (1-\theta)^{\beta-1}$$

$$= \underline{\text{Beta}(\alpha+3, \beta)}$$

$\alpha : \alpha + 3$

$\beta : \beta$

$$\text{mode} = \frac{\alpha+2}{\alpha+\beta+1}$$

$$\text{mean} = \frac{\alpha+3}{\alpha+\beta+3}$$

What is required to solve Linear reg questions:

- gradient ? $\nabla f(x, y, z) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right]$
- dot product $\langle v, u \rangle =$

$= v_1 u_1 + v_2 u_2 \dots v_n u_n$

- matrix multiplication

$A: m \times n \quad B: n \times l \rightarrow C: m \times l$

$AB \neq BA$ L2 norm

L2 norm : Euclidean norm

$$\|v\|_2 = \sqrt{v_1^2 + v_2^2 \dots v_n^2}$$

$$(X\theta)^T = \theta^T X^T$$

$$\underbrace{X^T X}_{} \theta^* = \dots \quad / \cdot (X^T X)^{-1}$$

$$\theta^* = (X^T X)^{-1} \dots$$