INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 10

Hassan Ashtiani

MAXIMUM LIKELIHOOD ESTIMATE

- Maximizes the probability of the observations given the parameters
  - $\boldsymbol{\alpha^{ML} = \underset{\alpha}{argmax}\, P\,(X|\alpha)}$
  - $\boldsymbol{\alpha^{ML} = \underset{\alpha}{argmin} - \sum_i \mathrm{LOG}\left(P\,(x^i|\alpha)\right)}$
- For bias of the coin
  - $\alpha^{ML} = \frac{n_0}{n_0+n_1}$

MAXIMUM A POSTERIORI ESTIMATE

- MAXIMIZES THE PROBABILITY OF THE PARAMETERS GIVEN THE OBSERVATIONS

  - $\boldsymbol{\alpha^{MAP} = \underset{\alpha}{argmax}\, P\,(\alpha|X)}$

$$\underset{\alpha}{argmax}\; p(\alpha|X) = \underset{\alpha}{argmax} \frac{p(X|\alpha)\,p(\alpha)}{p(X)} = \underset{\alpha}{argmax}\; p(X|\alpha)\,p(\alpha)$$

- $\alpha^{MAP} = \underset{\alpha}{argmin}\left(-\mathrm{LOG}(P(\alpha)) - \sum_{i=1}^{n} \mathrm{LOG}\,\boldsymbol{P}(x^i|\alpha)\right)$

$$\underset{\alpha}{\arg\max}\, p(\alpha|x) = \underset{\alpha}{\arg\max}\, p(x^i|\alpha)\, p(\alpha)$$

$$= \underset{\alpha}{\arg\max}\, \log p(x|\alpha)\, p(\alpha)$$

$$= \underset{\alpha}{\arg\max}\, \log p(\alpha) \prod_{i=1}^{n} p(x^i|\alpha)$$

$$= \underset{\alpha}{\arg\max} \left[\log p(\alpha) + \sum \log p(x^i|\alpha) \right]$$

$$= \underset{\alpha}{\arg\min} - \left[\log p(\alpha) + \sum \log p(x^i|\alpha) \right]$$

PRIOR VS POSTERIOR DISTRIBUTIONS

- $\mathrm{P}(\alpha)$ captures the **prior** distribution
- $\mathrm{P}(\alpha|X)$ captures the **posterior** distribution
- In other words,
  - We start by a prior belief about value of α
  - Our belief is updated after seeing some real data
  - This is a **Bayesian** Approach

$* \; P(\alpha) =$

MAP FOR COINS – UNIFORM PRIOR

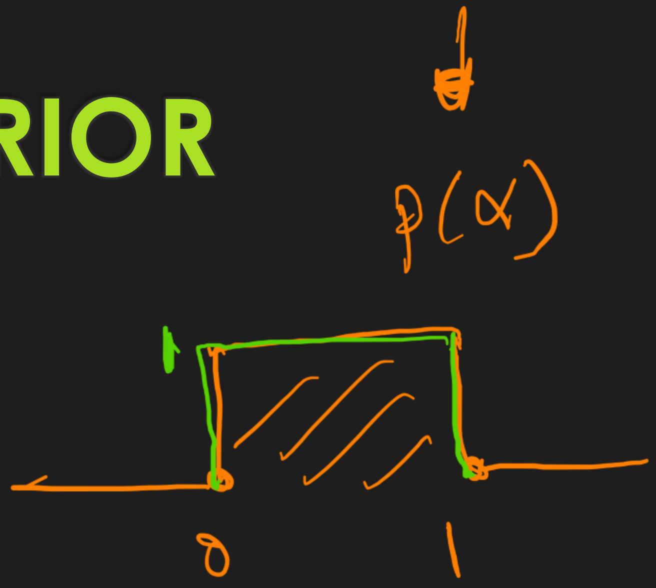
$x_1, \dots, x_n$: observations

$x_i \in \{0,1\}$, e.g. $0, 1, 0, 0, 1, 1, \dots$

$\alpha^{MAP} = \arg\max_{\alpha} p(\alpha | x) = ?$

$= \arg\max_{\alpha} p(x|\alpha)\, p(\alpha) = \arg\max_{\alpha} p(x|\alpha)$

same solution as ML.



#heads $= n_0$

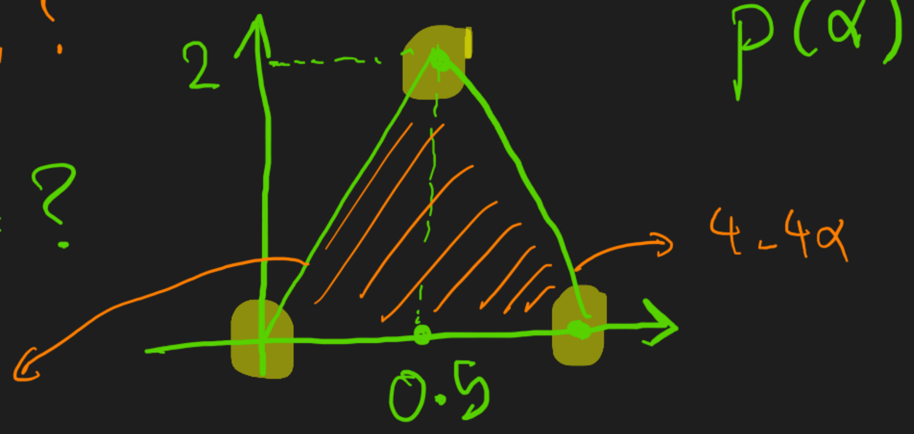
#tails $= n_1$

$n = n_0 + n_1$

MAP FOR COINS – NONUNIFORM PRIOR

Map solutions in terms of n_0, n_1?

$$\arg\max_{\alpha} \, p(x|\alpha)\,p(\alpha) = f(\alpha) \qquad \alpha^{MAP} = ?$$



4α

manually check $\alpha = 0, \alpha = 1, \alpha = 0.5$

$$\alpha \in (0, 0.5): \quad \frac{\partial f}{\partial \alpha} = \alpha^{n_0}(1-\alpha)^{n_1} 4\alpha = 0 \longrightarrow \alpha = \frac{n_0+1}{n_0+n_1+1}$$

$$\alpha \in (0.5, 1): \quad \frac{\partial f}{\partial \alpha} = \alpha^{n_0}(1-\alpha)^{n_1}(4-4\alpha) = 0 \longrightarrow \alpha = \frac{n_0}{n_0+n_1+1}$$

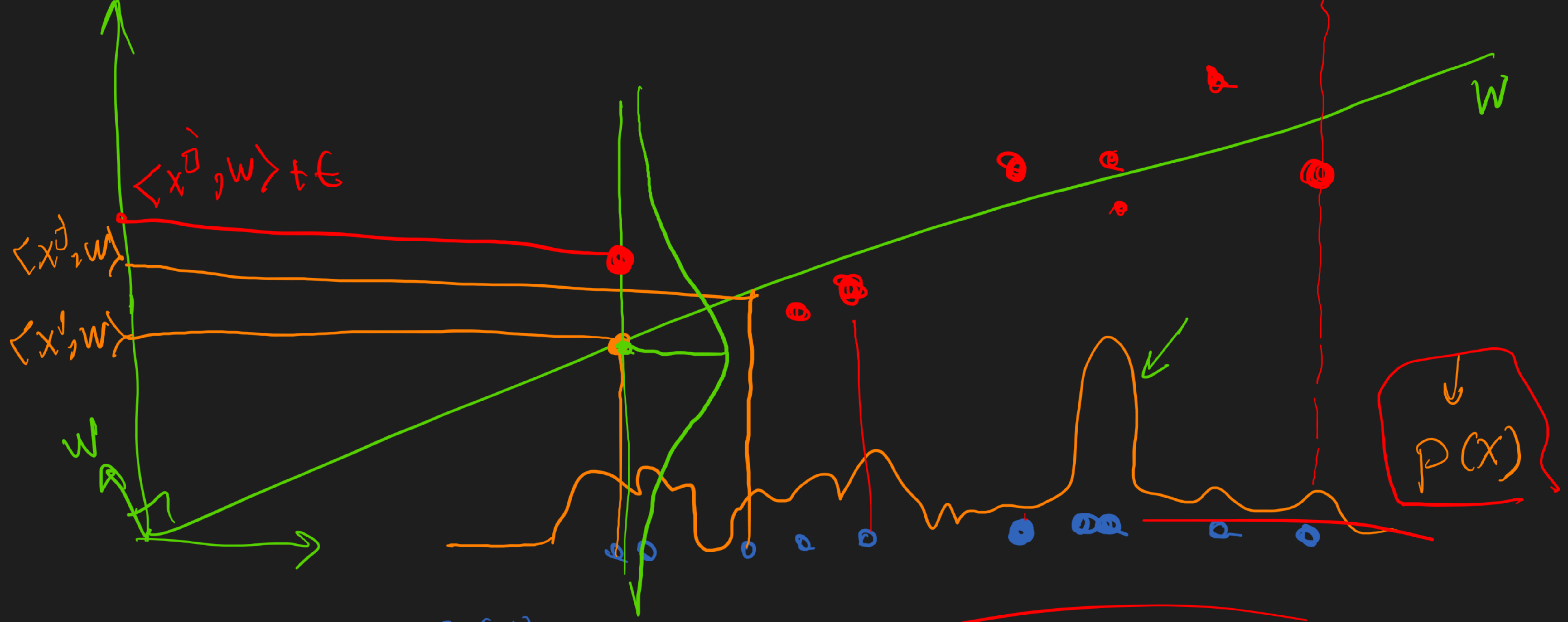
pick the best out of these 5.

$$\alpha^{MAP} = \begin{cases} \dfrac{n_0}{n_0+n_1+1} & n_0 > n_1 \\ \dfrac{n_0+1}{n_0+n_1+1} & n_1 > n_0 \\ \dfrac{n_0}{n_1+n_0} = 0.5 & n_0 = n_1 \end{cases}$$

A PROBABILISTIC MODEL FOR REGRESSION

- $P(y|x, W) = \left(\frac{1}{\gamma}\right) e^{-\frac{\left(x^T W - y\right)^2}{2\sigma^2}}$
- $P(x|W) = P(x)$

- ALTERNATIVELY, WE COULD SAY THAT
  - $Y = X^T W + \epsilon$ WHERE $\epsilon \sim \mathrm{N}(0,1)$



$$X \sim p(X)$$

$$\langle x, w \rangle$$

$$\langle x, w \rangle + \epsilon$$

- $P(y|x,W) = \left(\frac{1}{\gamma}\right) e^{-\frac{\left(x^T W - y\right)^2}{2\sigma^2}}$
- $P(x|W) = P(x)$

$\hat{y} = w^T x$

- INPUT: $((x^1, y^1), \ldots, (x^n, y^n))$ I.I.D. SAMPLE
- EACH (x^i, y^i) IS DRAWN ACCORDING TO P(X,Y) = P(x)P(Y|X)

need to know W

- WHAT DO WE NEED TO BE ABLE TO
  - MAKE PREDICTIONS? → just need W
  - GENERATE MORE DATA? W and P(x)
- FIND MAXIMUM LIKELIHOOD ESTIMATE FOR W? → we don't estimate

ML ESTIMATE

$$Z = \{(x^i, y^i)\}_{i=1}^{n}$$

$$\underset{W}{\operatorname{argmax}}\ p(Z \mid W) = \underset{W}{\operatorname{argmax}} \prod p(x^i, y^i \mid w)$$

$\leq a$ ---- ϵ

MAXIMUM A POSTERIORI AND RLS

- $P(y|x, W) = \left(\frac{1}{\gamma}\right) e^{-\frac{\left(x^T W - y\right)^2}{2\sigma^2}}$
- $P(x|W) = P(x)$
- $P(W) = \frac{1}{\beta} e^{-\lambda \|W\|_2^2}$

MAP ESTIMATE