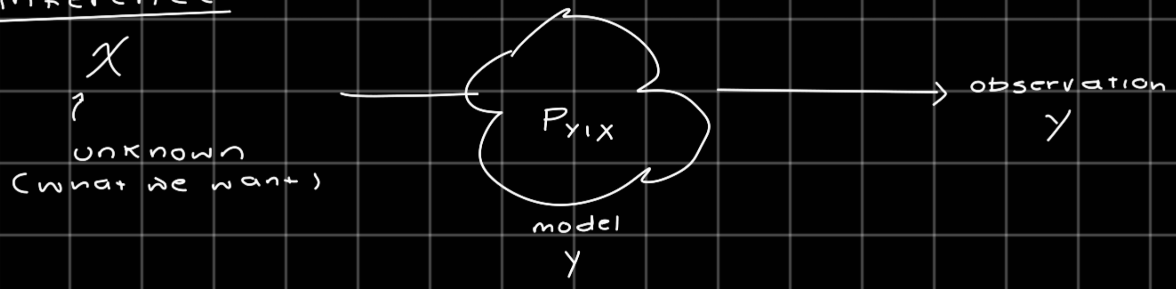
inference



↳ one approach: maximize likelihood of observation Y under our model

$\text{MLE}(X|Y)$ = "maximum likelihood estimate of $\mathcal{X}$ given our observation Y"

$$:= \underset{x}{\text{argmax}}\ P_{Y|X}(Y|x)$$

ex: $Y \sim \mathcal{N}(x, \sigma^2)$

$$\text{MLE}(X|Y) = \underset{x}{\text{argmax}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y-x)^2}{2\sigma^2}}$$

$$= Y$$

Ex: $\underbrace{Y_1, Y_2, \ldots, Y_n}_{Y} \underset{iid}{\sim} \mathcal{N}(X, \sigma^2)$

$$\text{MLE}(X|Y) = \underset{X}{\text{argmax}} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{\sum_{i=1}^{n}(Y_i - x)^2}{2\sigma^2}}$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} Y_i}_{\text{emperical mean of observations}}$$

Ex: $Y \sim \mathcal{N}(0, X)$

$$\text{MLE}(X|Y) = \underset{x}{\text{argmax}} \frac{1}{\sqrt{2\pi x}} e^{-Y^2/2x}$$

maximizing the likelihood is equavelent to maximizing log of likelihood

$$= \underset{x}{\text{argmax}} -\frac{1}{2}\log 2\pi x - \frac{y^2}{2x}$$

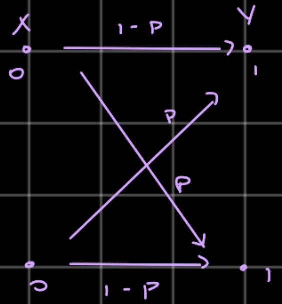
$$= \underset{x}{\text{argmax}} -\log x - \frac{y^2}{x}$$

$$= \underset{x}{\text{argmin}} \underbrace{\log x + \frac{y^2}{x}}_{-\Lambda(x)}$$

$$\frac{\partial}{\partial x} \ell(x) = \frac{1}{x} - \frac{y^2}{x^2}$$

$$= y^2$$

Now, let's take an example where we have a prior.



$$P_{Y|X}(y|x) = \begin{cases} 1-p & y = x \\ p & y \neq x \end{cases}$$

$$MLE(X|Y) = \begin{cases} y & p < \frac{1}{2} \\ 1-y & p \geq \frac{1}{2} \end{cases}$$

what if we know:

$$\pi_0 = P(X=0) \quad \& \quad \pi_1 = P(X=1) \quad \Big\} \text{ a prior!}$$

since we have prior π, we can compute a posteriori probabilities

$P_{X|Y}(x|y)$ $\forall y$.

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)\pi_x}{\underbrace{P_{Y|X}(y|0)\pi_0 + P_{Y|X}(y|1)\pi_1}_{= P_Y(y)}}$$

Maximum a posteriori estimate:

$$MAP(X|Y) = \arg\max_x P_{X|Y}(x|Y) = \arg\max_x P_{Y|X}(Y|x)\pi_x$$

puts more importance on certain likelihoods than other likelihoods

Ex:

X	Y	$P_{Y\|X}(y\|x)\pi_x$
0	0	$(1-p)\pi_0$
1	0	$p\pi_1$
0	1	$p\pi_0$
1	1	$(1-p)\pi_1$

$$MAP(X|0) = \begin{cases} 0 & \text{if } \underbrace{(1-p)\pi_0 > p\pi_1}_{\pi_0 > p} \\ 1 & \text{if } \pi_0 \leq p \end{cases}$$

$$MAP(X|1) = \begin{cases} 0 & \text{if } p > \pi_1 \\ 1 & \text{if } p \leq \pi_1 \end{cases}$$

Note: if $\pi_x = \frac{1}{2}$, MLE = MAP

Both MLE & MAP from previous example can be passed as special cases of "Threshold tests":

① Binary hypothesis testing:

we have 2 "hypotheses" $H_0 = (X = 0)$ "null hypothesis"

$H_1 = (X = 1)$ "alternative hypothesis"

Likelihood ratio $L(y) := \dfrac{P_{Y|X}(y \mid 1)}{P_{Y|X}(y \mid 0)} = \dfrac{\text{likelihood of observing } y \text{ under } H_1}{\text{likelihood of observing } y \text{ under } H_0}$

Previous example:

$$\mathrm{MLE}(X \mid Y) = \begin{cases} 1 & \text{if } L(y) \geq 1 \\ 0 & \text{if } L(y) < 1 \end{cases}$$

$\Leftrightarrow P(y \mid 1) \geq P(y \mid 0)$

if $y = 0$ $p \geq 1 - p$ $\Leftrightarrow p \geq \frac{1}{2}$

if $y = 1$ $(1-p) \geq p$ $\Leftrightarrow p \leq \frac{1}{2}$

$$\mathrm{MAP}(X \mid Y) = \begin{cases} 1 & \text{if } L(y) \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{if } L(y) < \frac{\pi_0}{\pi_1} \end{cases}$$

So, threshold tests w/ threshold η is formulated as:

$$\hat{X}(y) = \begin{cases} 1 & \text{if } L(y) \geq \eta \\ 0 & \text{if } L(y) < \eta \end{cases}$$

Special cases:

MLE: $\eta = 1$

MAP: $\eta = \frac{\pi_0}{\pi_1}$ (when prior known)

ex: $H_0 : Y \sim \mathcal{N}(0, \sigma^2)$

$H_1 : Y \sim \mathcal{N}(1, \sigma^2)$

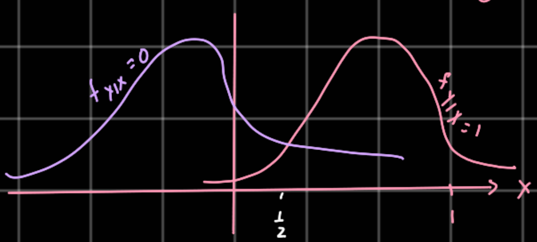
Q/ what's threshold test look like?

$$L(y) = \frac{f_{Y|X}(y \mid 1)}{f_{Y|X}(y \mid 0)} = \frac{e^{-(y-1)^2/2\sigma^2}}{e^{-y^2/2\sigma^2}} = e^{\frac{1}{2\sigma^2}(y^2 - y^2 + 2y - 1)}$$

$$= e^{\frac{1}{\sigma^2}(y - \frac{1}{2})}$$

$$L(y) \geq \eta \Rightarrow e^{\frac{1}{\sigma^2}(y - \frac{1}{2})} \geq \eta$$

$$\Leftrightarrow y \geq \frac{1}{2} + \sigma^2 \log \eta$$



$$\text{MLE}(X|Y) = \begin{cases} 1 & \text{if } Y \geq \frac{1}{2} \\ 0 & \text{o/w} \end{cases}$$

$$\text{MAP}(X|Y) = \begin{cases} 1 & Y \geq \frac{1}{2} + \sigma^2 \log \frac{\pi_0}{\pi_1} \\ 0 & \text{o/w} \end{cases}$$

$$\arg\max_x \; P_{Y|X}(y|x)\,\pi_x$$

$$x = 1 \iff P_{Y|X}(y|1)\,\pi_1 \geq P_{Y|X}(y|0)\,\pi_0$$

$$\iff L(y) \geq \frac{\pi_0}{\pi_1}$$

Binary hypothesis testing

$H_0 : Y \sim P_{Y|X=0}$
$H_1 : Y \sim P_{Y|X=1}$

we want to discriminate between these 2 hypotheses given observation Y

These tests seem like a pretty good class of tests

Q: Are these tests optimal in some sense?

To define optimal we need to take careful look at the problem:

A decision rule "test" is just a function

$\hat{X} : \mathcal{Y} \to \{0, 1\}$

input = observation

output = hypothesis

For any test $\hat{X}$, there are 2 fundamental types of errors:

1. Type I: False positive $\hat{X}(Y) = 1$ but $X = 0$
2. Type II: False negative $\hat{X}(Y) = 0$ but $X = 1$

False positive error rate: $P(\hat{X}(Y) = 1 \mid X = 0)$ — fun of the likelihoods $P_{Y|X}(\cdot|0)$

False negative error rate: $P(\hat{X}(Y) = 0 \mid X = 1)$ — fcn of $P_{Y|X}(\cdot|1)$

Basic idea: type 1 & 2 error rates are in tension w/ each other