PROBLEM 1B

To prove: $P(Y|X) = \frac{1}{1+e^{-y(\beta_0+\beta_1 x)}}$ if $Y \in \{-1, 1\}$

Given: $P(Y=+1|X) = S(\beta_0+\beta_1 X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$

$$\Rightarrow P(Y=-1|X) = 1 - S(\beta_0+\beta_1 X) = \frac{1+e^{-(\beta_0+\beta_1 x)} - 1}{1+e^{-(\beta_0+\beta_1 x)}}$$

$$= \frac{e^{-(\beta_0+\beta_1 x)}}{e^{-(\beta_0+\beta_1 x)}\left[e^{(\beta_0+\beta_1 x)}+1\right]} = \frac{1}{1+e^{(\beta_0+\beta_1 x)}}$$

Thus, to generalize, we use Y as a coefficient in the exponential term

Hence Proved.

To prove: Log likelihood for m data points can be:

$$\ln \mathcal{L}(\beta_0, \beta_1) = -\sum_{i=1}^{m} \ln\left(1+e^{-y_i(\beta_0+\beta_1 x_i)}\right)$$

We know from above that $\mathcal{L}(\beta_0, \beta_1) = \prod_{i=1}^{m} \frac{1}{1+e^{-y(\beta_0+\beta_1 x)}}$

$$\Rightarrow \ln(\mathcal{L}(\beta_0, \beta_1)) = \sum_{i=1}^{m} \frac{1}{1+e^{-y(\beta_0+\beta_1 x)}} = \sum_{i=1}^{m}\left(\ln(1) - \ln\left(1+e^{-y(\beta_0+\beta_1 x)}\right)\right)$$

$$= -\sum_{i=1}^{m} \ln\left(1+e^{-y(\beta_0+\beta_1 x)}\right)$$

Hence Proved.

3 A] Given: $P(X=x, Y=y) = P(Y=y)\,P(X=x \mid Y=y)$

Distribution support $x \in \mathbb{R}, y \in \{-1,+1\}$

$$P(Y=+1) = P(Y=-1) = \frac{1}{2}$$

$$P(X=x \mid Y=+1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2}}$$

$$P(X=x \mid Y=-1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+5)^2}{2}}$$

Prove $P(X=x \,\&\, Y=y) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+5y)^2}{2}}$ ⇒ This is easy just multiply $P(Y=y)$ for both $+1$ & -1 cases

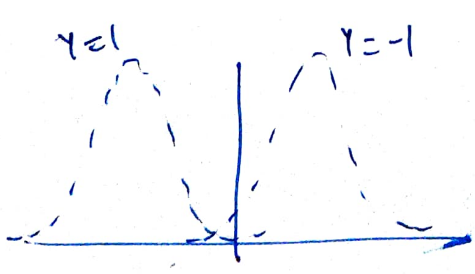
3 B] $P(X=x \mid Y=+1)$

$P(X=x \mid Y=-1)$ in a single figure

3C] Write Bayes optimal classifier $h^*(x)$ given P

⇒ classify x based on a threshold

⇒ this threshold can be 0 because each $Y=1$ & -1 is unimodal & symmetric peaks w/ $X=0$ as axis of symmetry



Optimal classifier

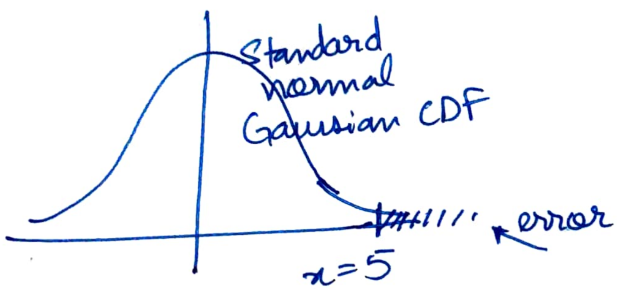
$$x = \begin{cases} < 0 & y = 1 \\ > 0 & y = -1 \end{cases}$$

3d) compute error of Bayes optimal classifier

$$\Pr_{(x,y)\sim P} (h^*(x) \neq y) = \mathbb{E}_{(x,y)\sim P} \left[\mathbb{1}_{h^*(x) \neq y} \right]$$

result should be of the form $1 - \Phi(c)$ where Φ is Gaussian CDF

$$\boxed{1 - \Phi(5)}$$



3e] $p(Y = +1 \mid X = x) = \dfrac{1}{1 + e^{-\beta_0 - \beta_1 x}}$

show that the given distribution satisfies this assumption

$$P(Y = +1 \mid X = x) = \frac{P(Y = +1 \ \& \ X = x)}{P(X = x)} = \frac{P(X = x \mid Y = +1)\, P(Y = +1)}{P(X = x)}$$

$$P(X = x) = P(X = x \mid Y = +1) \cdot P(Y = +1) + P(X = x \mid Y = -1) \cdot P(Y = -1)$$

$$= \frac{1}{2} \left(P(X = x \mid Y = 1) + P(X = x \mid Y = -1) \right)$$

$$a = x - 5$$

$$\cancel{1 - a = -x + 5}$$

$$10 + a = x + 5$$

$$P(Y=+1 \mid X=x) = \frac{\cancel{\frac{1}{2}} \cdot \cancel{\frac{1}{\sqrt{2\pi}}}\, e^{\frac{-(x-5)^2}{2}}}{\cancel{\frac{1}{2}}\left(\cancel{\frac{1}{\sqrt{2\pi}}}\, e^{-\frac{(x-5)^2}{2}} + \cancel{\frac{1}{\sqrt{2\pi}}}\, \underbrace{e^{-\frac{(x+5)^2}{2}}}_{y}\right)} \cancel{\Rightarrow e^{\frac{-a^2}{2}}}$$

$$= \frac{1}{1 + \dfrac{e^{-\frac{(x+5)^2}{2}}}{e^{-\frac{(x-5)^2}{2}}}}$$

$$= \frac{1}{1 + e^{-\frac{1}{2}\left(\cancel{x^2} + \cancel{25} + 10x - \cancel{x^2} - \cancel{25} + 10x\right)}}$$

$$\frac{1}{1 + e^{-\beta_0 - \beta_1 x}} = \frac{1}{1 + e^{-10x}}$$

$$\Rightarrow \boxed{\beta_0 = 0, \quad \beta_1 = 10}$$

4) $$\min_{c_1,\ldots,c_k} \sum_{i=1}^{n} \|x_i - c(x_i)\|_2^2$$

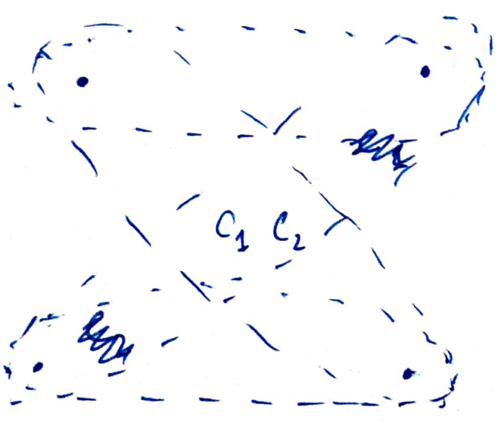
show k-means does not always find optimal objective

Given: OPT is optimal objective

For every $t > 1$, show $\exists$ an instance of the above optimization problem for which k-means might find a solution whose objective value $\geq t \cdot OPT$

i.e. Find $x_1, \ldots, x_n$ for which k-means with <u>some</u> bad initialization of centers will output a set of centers that achieves an objective of $t \cdot OPT$

Start w/ example of 4 points in a 2D plane, 2 clusters; generalize this example to p-dimensions, n data points, k clusters



Clusters
Always stuck crossing diagonal
& keep switching main/off diagonal