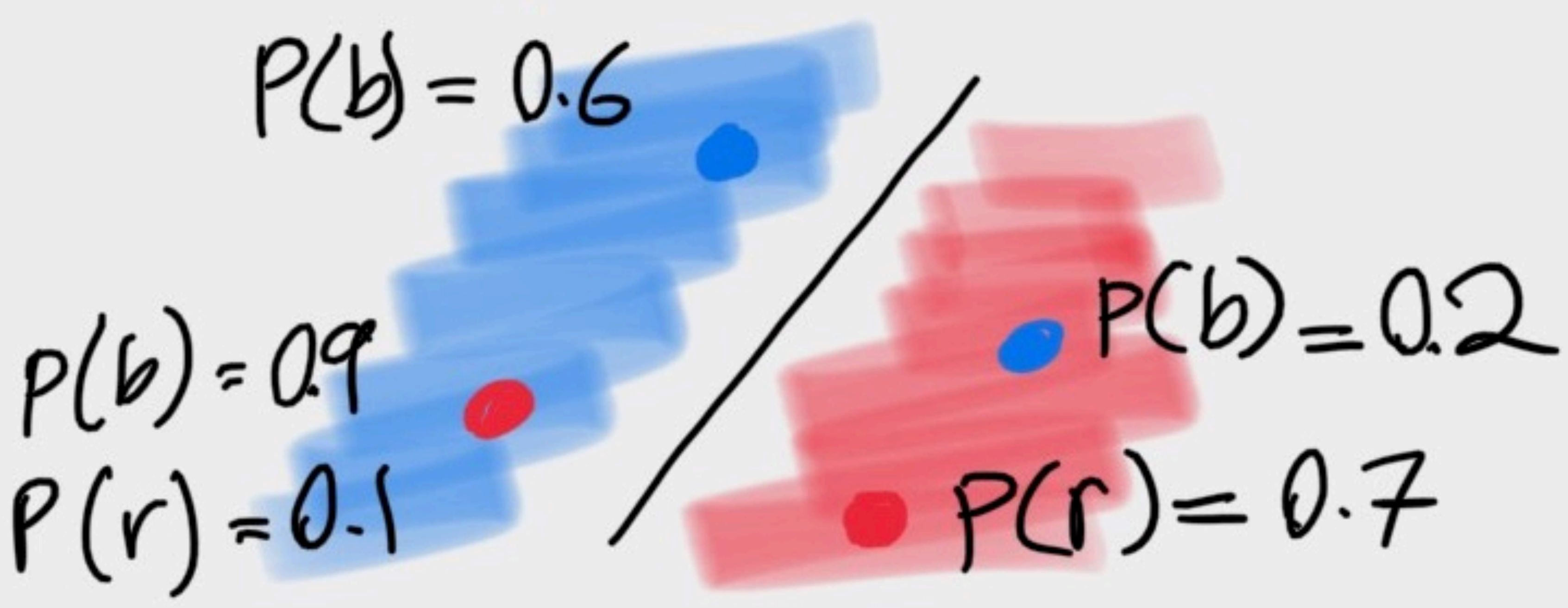
Maximum Likelihood

Pick the model that gives the existing labels the highest probability.
So we try to maximize this probability.

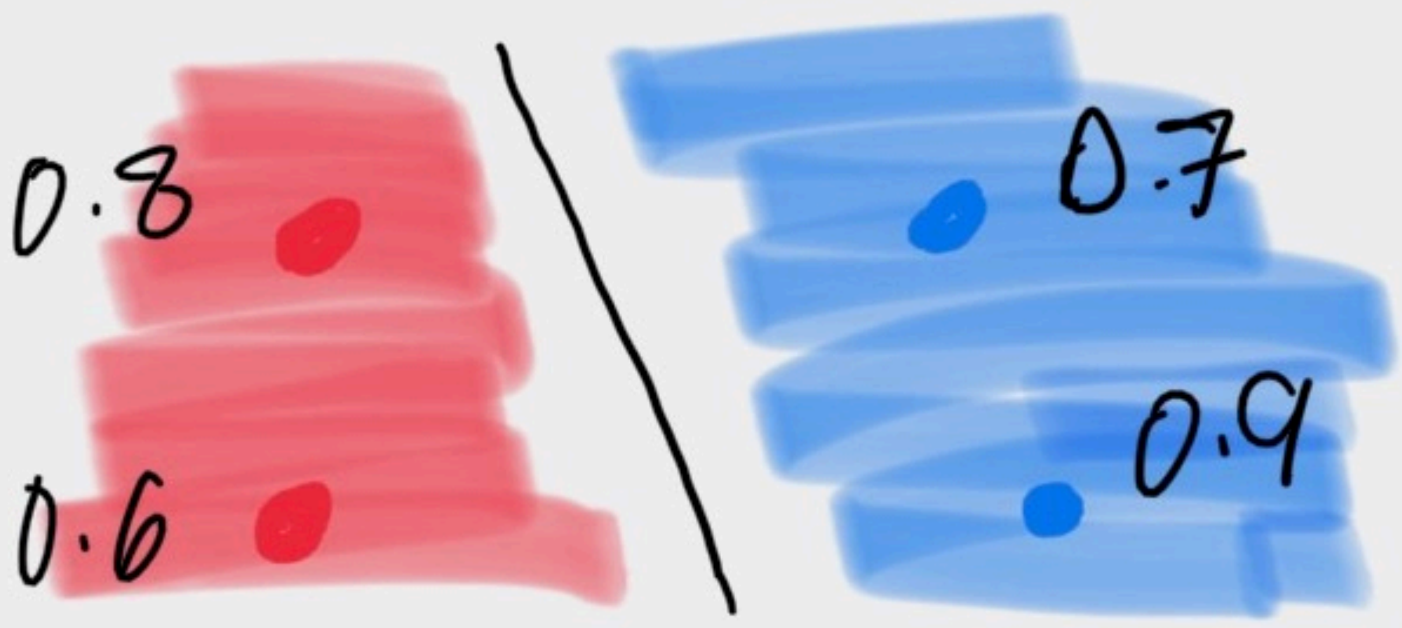
example:



$\hat{y} = \sigma(Wx+b) = P(blue)$

We assume the points are iid:

$$P(all) = \underbrace{0.1 \times 0.7}_{\text{red dots}} \times \underbrace{0.6 \times 0.2}_{\text{blue dots}} = \underset{\text{small}}{\underset{\downarrow}{0.0084}}$$



$\boxed{P(all)} = \underbrace{0.3024}$ better than the other model.

Maximum likelihood tries to maximize this.

When dealing with maximum likelihood we prefer sums to products. $\longrightarrow \log_e() == \ln()$

$\ln(x) < 0$ if $x \in [0,1]$. So we use $\boxed{-\ln(x)}$.

$\underbrace{\text{Cross Entropy}}_{\text{we can think of it as error}}$ of good models is smaller.

Error Function = Cross Entropy

So we now change the goal to minimizing cross entropy.

Probabilities –

	G	R	B
P(gift)	0.8	0.7	0.1
P(no gift)	0.2	0.3	0.9

most likely event? choosing the largest prob. in each column.

$0.8 \times 0.7 \times 0.9$

$$\text{Cross Entropy} = -\sum_{i=1}^{m} y_i \ln(P_i) + (1-y_i)\ln(1-P_i)$$

it can tell us how close two vectors are. e.g.,

$CE[(1,1,0), (0.8, 0.7, 0.1)] = \boxed{0.69}$

$CE[(0,0,1), (0.8, 0.7, 0.1)] = 5.12$

likely to happen

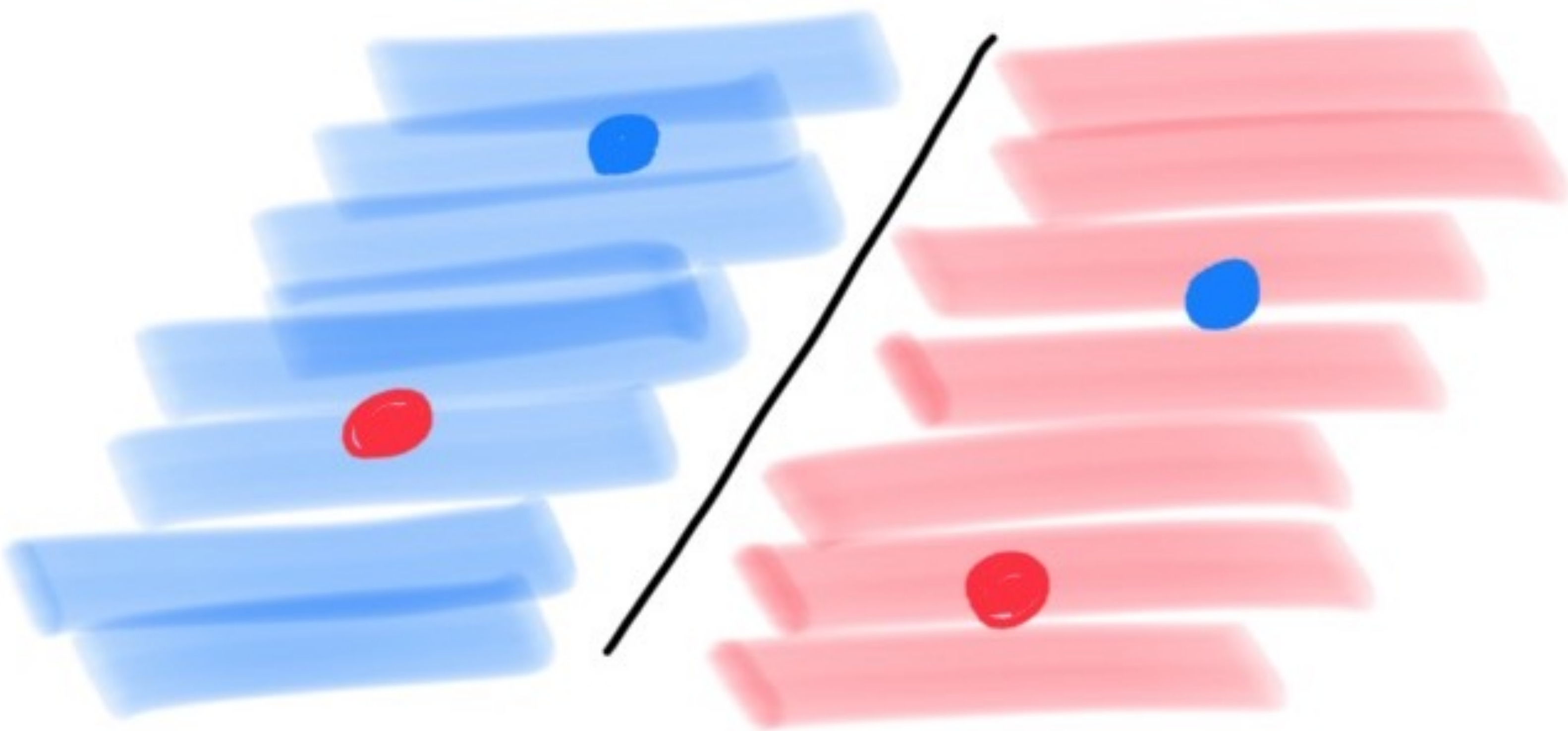
Multi-class Entropy:

number of classes

probability of i and j

$$CE = -\sum_{i=1}^{n} \sum_{j=1}^{\boxed{m}} y_{ij} \ln(p_{ij})$$

0/1 to indicate which event we are talking about.



$$\begin{cases} \text{if } y=1 \longrightarrow P(\text{blue}) = \hat{y} \longrightarrow \text{Error} = -\ln \hat{y} \\ \text{if } y=0 \longrightarrow P(\text{red}) = 1-P(\text{blue}) = 1-\hat{y} \longrightarrow \text{Error} = -\ln(1-\hat{y}) \end{cases}$$

Error function: $\frac{-1}{m} \sum_{i=1}^{m} (1-y_i)(\ln(1-\hat{y}_i)) + y_i \ln \hat{y}_i$

binary classification:

Same as:

$$E(W,b) = \frac{-1}{m} \sum_{i=1}^{m} (1-y_i) \ln(1-\sigma(Wx^i+b)) + y_i \ln(\sigma(Wx^i+b))$$

Multi-Class classification:

$$\frac{-1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} \ln \hat{y}_{ij}$$