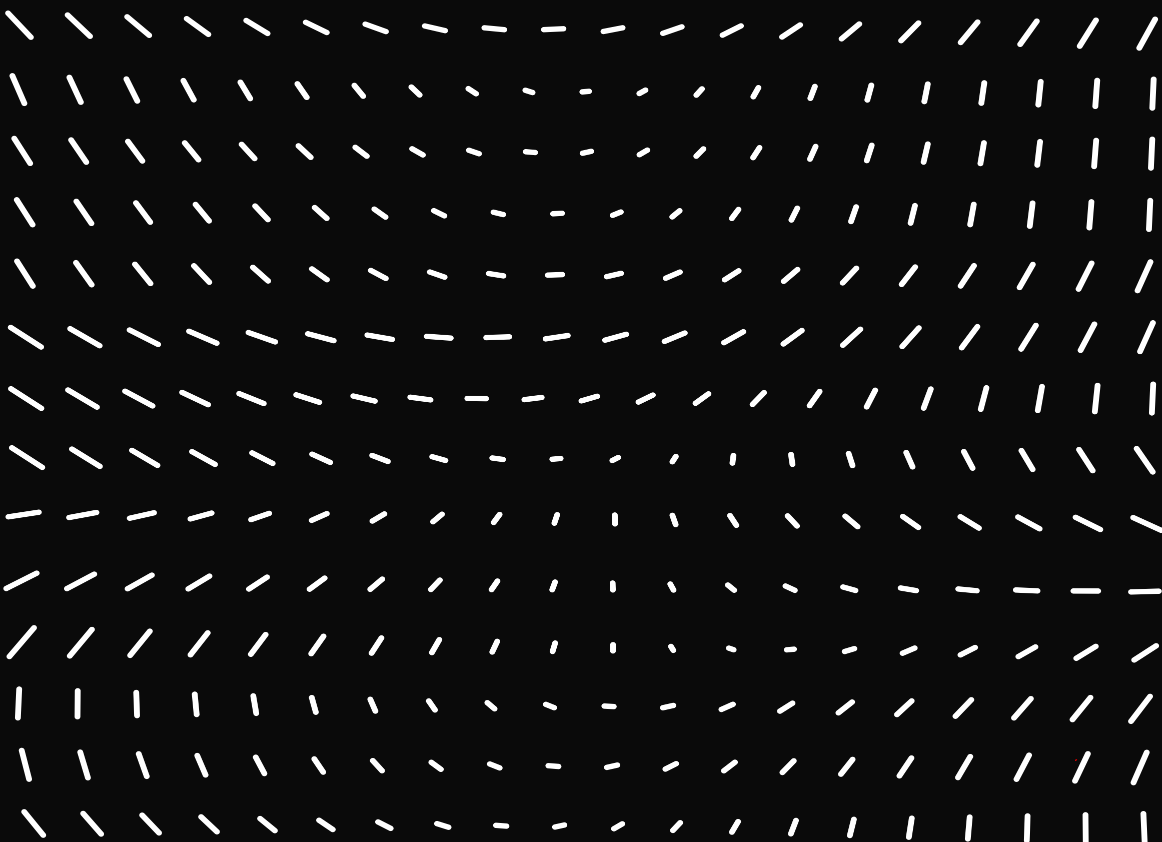


# Ceng 462 — Artificial Intelligence

## Naive Bayes:

### Classification:

ex: Spam filter

setup → two labels — decisions: "spam" or "ham"

Features: words {"FREE" ... }
text patterns
Nontext → Sender is in Contacts or Not
wide Broadcast.

↳ * classification = given inputs x, predict labels (classes) y

ex: Simple Digit Recognition

* Feature values are binary True or False.

$Y = 10 \quad \{0, 1, \dots 9\}$

$|F|^n \rightarrow \{0, 1\}$

total number of param is linear in n

| prior prob | label |
|---|---|
| 0.1 | 1 |
| 0.1 | 2 |
| 0.1 | 3 |
| 0.1 | 4 |
| ⋮ | ⋮ |

↳ Inference Method

↳ Estimates of local conditional probability tables:

$P(Y)$ → Prior probability
$P(F|Y)$ ↳ Feature probabilities
updates $\xrightarrow{\text{gives}}$ labels

## Parameter Estimation:

$$P_{\text{maximum likelihood}} = \frac{\text{count}(x)}{\text{total samples}}$$

smoothing → if count comes up as 0 during the training,

$$P_{LAP}(x) = \frac{\text{count}(x)+1}{\sum_x [\text{count}(x)+1]} = \frac{c(x)+1}{N+1}$$

if 1 is k   k is Laplace's estimate.

$$P_{LAP,\,100}(x) = \left\langle \frac{102}{203}, \frac{101}{203} \right\rangle$$

$+100$

## Perceptroons:

# CENG 462 - AI

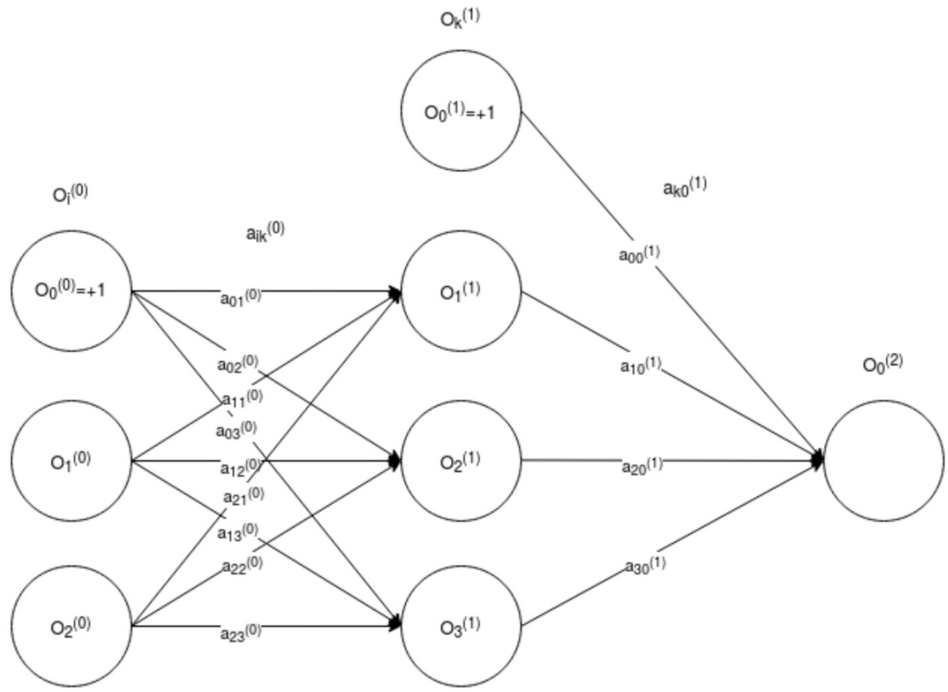
## Make-Up Homework

Arda Karaman
2237568

### Regression:

Ground truth $y = [x_1, x_2, x_3]$

$$\boxed{\begin{aligned} O_0^{(0)} &= 1 \\ O_1^{(0)} &= x_1 \\ O_2^{(0)} &= x_2 \end{aligned}}$$




### For the first feed:

$a_{ij}$ has initial random values.

$$O_1^{(1)} = 1 \cdot a_{01}^{(0)} + x_1 \cdot a_{11}^{(0)} + x_2\, a_{21}^{(0)}$$

$$O_1^{(1)'} = \sigma\left(O_1^{(1)}\right) = \frac{1}{1+e^{-O_1^{(1)}}}$$

$$O_2^{(1)} = 1 \cdot a_{02}^{(0)} + x_1 \cdot a_{12}^{(0)} + x_2\, a_{22}^{(0)}$$

$$O_2^{(1)'} = \sigma\left(O_2^{(1)}\right) = \frac{1}{1+e^{-O_2^{(1)}}}$$

$$O_3^{(1)} = 1 \cdot a_{03}^{(0)} + x_1 \cdot a_{13}^{(0)} + x_2\, a_{23}^{(0)}$$

$$O_3^{(1)'} = \sigma\left(O_3^{(1)}\right) = \frac{1}{1+e^{-O_3^{(1)}}}$$

Hidden Layer Values

$$O_0^{(2)} = O_0^{(1)} \cdot a_{00}^{(1)} + O_1^{(1)'} \cdot a_{10}^{(1)} + O_2^{(1)'} \cdot a_{20}^{(1)} + O_3^{(1)'} \cdot a_{30}^{(1)}$$

(out) $O_0^{(2)'} = \sigma\left(O_0^{(2)}\right)$

$$ {}^{(out)}O_0^{(2)'} = \frac{1}{1+e^{-O_0^{(2)}}} $$

## Calculating the error:

↳ we took ground truth value as $y$. learning rate as $\alpha$.

$$ \text{Error} = SSE(y, y') = (y - y')^2 $$

$$ \text{Error} = SSE(y, {}^{(out)}O_0^{(2)}) = (y - {}^{(out)}O_0^{(2)})^2 $$

## Backpropagation Algorithm:

We use error value to update the weights $a_{ij}^{(k)}$

For $a_{10}^{(1)}$:

$$ \frac{d(\text{Error})}{d(a_{10}^{(1)})} = \frac{d(\text{Error})}{d({}^{(out)}O_0^{(2)})} \cdot \frac{d({}^{(out)}O_0^{(2)})}{d(O_0^{(2)})} \cdot \frac{d(O_0^{(2)})}{d(a_{10}^{(1)})} $$

$$ \frac{d(\text{Error})}{d({}^{(out)}O_0^{(2)})} = -2 \cdot (y - {}^{(out)}O_0^{(2)}) $$

$$ \frac{d({}^{(out)}O_0^{(2)})}{d(O_0^{(2)})} = \frac{d(\sigma(O_0^{(2)}))}{d(O_0^{(2)})} = {}^{(out)}O_0^{(2)} \cdot (1 - {}^{(out)}O_0^{(2)}) $$

$$ \frac{d(O_0^{(2)})}{d(a_{10}^{(1)})} = O_1^{(1)'} \cdot (a_{10}^{(1)})^{(1-1)} = O_1^{(1)'} $$

$$ \frac{d(\text{Error})}{d(a_{10}^{(1)})} = \text{change in } a_{10}^{(1)} = \underbrace{\left[-2 \cdot (y - {}^{(out)}(O_0^{(2)}))\right] \cdot \left[{}^{(out)}O_0^{(2)} \cdot (1 - {}^{(out)}O_0^{(2)})\right] \cdot O_1^{(1)'}} $$

The new weight value of $a_{10}^{(1)}$ $= \quad a_{10}^{(1)} = a_{10}^{(1)} - \alpha \cdot \dfrac{d(\text{Error})}{d(a_{10}^{(1)})}$

* This procedure is the same for all $a_{ij}^{(1)}$.

For $a_{11}^{(0)}$ $=$ $\dfrac{d(\text{Error})}{d(a_{11}^{(0)})} = \dfrac{d(\text{Error})}{d({}^{(out)}O_0^{(2)})} \cdot \dfrac{d({}^{(out)}O_0^{(2)})}{d(O_0^{(2)})} \cdot \dfrac{d(O_0^{(2)})}{d(O_1^{(1)'})} \cdot \dfrac{d(O_1^{(1)'})}{d(O_1^{(1)})} \cdot \dfrac{d(O_1^{(1)})}{d(a_{11}^{(0)})}$

$$\frac{d(\text{Error})}{d({}^{(out)}O_0^{(2)})} = -2\cdot(y - {}^{(out)}O_0^{(2)})$$

$$\frac{d({}^{(out)}O_0^{(2)})}{d(O_0^{(2)})} = \frac{d(\sigma(O_0^{(2)}))}{d(O_0^{(2)})} = {}^{(out)}O_0^{(2)}\cdot(1 - {}^{(out)}O_0^{(2)})$$

$$\frac{d(O_0^{(2)})}{d(O_1^{(1)'})} = \left(O_1^{(1)'}\right)^{(1-1)}\cdot a_{10}^{(1)} = a_{10}^{(1)}$$

$$\frac{d(O_1^{(1)'})}{d(O_1^{(1)})} = \frac{d(\sigma(O_1^{(1)}))}{d(O_1^{(1)})} = O_1^{(1)}(1 - O_1^{(1)})$$

$$\frac{d(O_1^{(1)})}{d(a_{11}^{(0)})} = \cancelto{0}{1\cdot a_{01}^{(0)}} + x_1\cdot(a_{11}^{(0)})^{(1-1)} + \cancelto{0}{x_2\, a_{21}^{(0)}} = x_1$$

- $\underbrace{\dfrac{d(\text{Error})}{d(a_{11}^{(0)})}}_{\text{change in } a_{11}^{(0)}} = \underbrace{\left[-2\cdot(y - {}^{(out)}O_0^{(2)})\right]\cdot\left[{}^{(out)}O_0^{(2)}\cdot(1 - {}^{(out)}O_0^{(2)})\right]\cdot\left[a_{10}^{(1)}\right]\cdot\left[O_1^{(1)}\cdot(1 - O_1^{(1)})\right]\cdot\left[x_1\right]}$

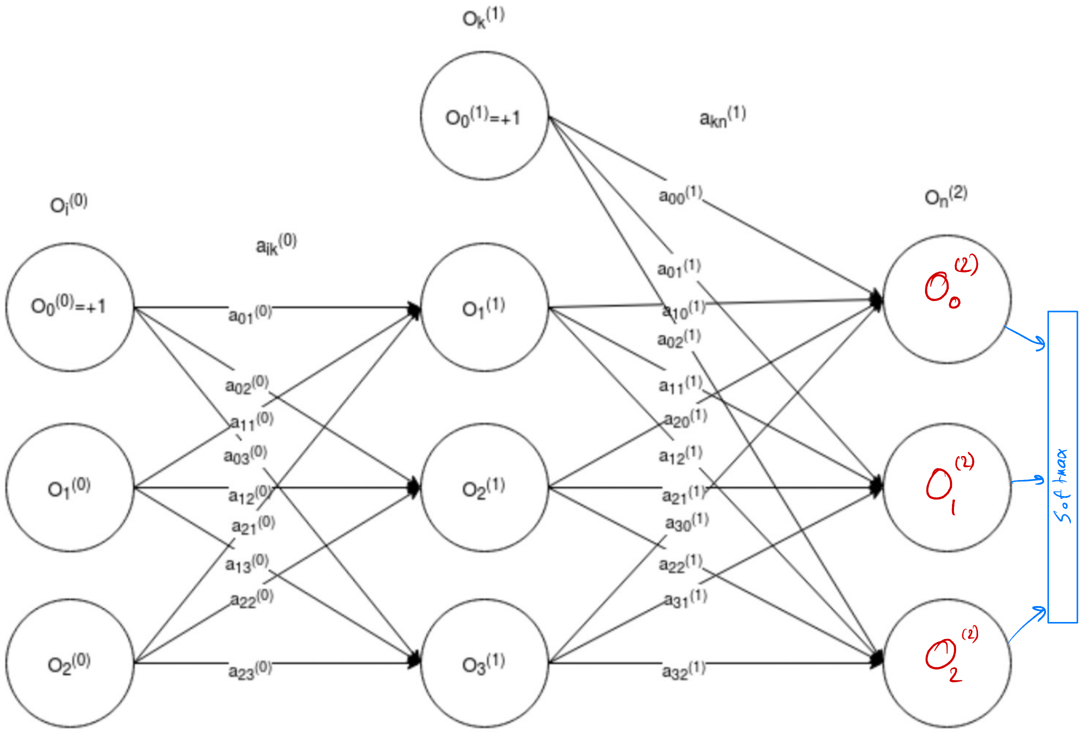
The new weight for $a_{11}^{(0)}$ $=$ $a_{11}^{(0)} = a_{11}^{(0)} - \alpha\cdot\dfrac{d(\text{Error})}{d(a_{11}^{(0)})}$

* This procedure is the same for all $a_{ij}^{(0)}$.

# Classification:

The ground truth:
$[l_0, l_1, l_2]$

$$\boxed{\begin{aligned} O_0^{(0)} &= 1 \\ O_1^{(0)} &= x_1 \\ O_2^{(0)} &= x_2 \end{aligned}}$$




✱ The only difference between classification and regression is that we calculate the error in classification as:

$$\underline{\text{Error:}} \quad CE(l, l') = -\sum_i l_i \cdot \log(l_i')$$

$$\text{so for } O_0^{(2)} \rightarrow CE\left(l_0, O_0^{(2)}\right) = l_0 \cdot \log\left(O_0^{(2)}\right)$$

$\Rightarrow$ Total Error $= CE\left(l = [l_0, l_1, l_2],\ l' = [O_0^{(2)}, O_1^{(2)}, O_2^{(2)}]\right) = \sum_{n=0}^{2} l_n \cdot \log\left(O_n^{(2)}\right)$

- So for updating weight $a_{10}^{(1)}$:

$$\frac{d(\text{total Err})}{d(a_{10}^{(1)})} = \frac{d(\text{total Err})}{d(O_0^{(2)})} \cdot \frac{d(O_0^{(2)})}{d(a_{10}^{(1)})}$$

$$\hookrightarrow \frac{d(\text{total Err})}{d(O_0^{(2)})} = \left[l_0 \cdot \log(O_0^{(2)}) + \cancel{l_1 \cdot \log(O_1^{(2)})} + \cancel{l_2 \cdot \log(O_2^{(2)})}\right] \big/ d(O_0^{(2)})$$

$$= l_0 \cdot \frac{1}{O_0^{(2)} \cdot \ln 10}$$

$\Rightarrow$ The rest is same as above.

$0.5 \cdot 0.75 = 0.375$

$0.5 \cdot 0.6 = 0.3$