Feature vector $x = [x_1, x_2, \dots, x_K]$ → does doc. contain 'like'? → 0 / → 1

Param. vector $\underline{w} = [w_1, w_2, \dots, w_K]$ and b a scalar (bias).

Compute $z = w^T x + b = \sum_{k=1}^{K} w_k x_k + b = (w_1 x_1 + \dots + w_K x_K) + b$

$$\boxed{p(+|x) = \sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(w^T x + b)}}} > 0.5? \begin{cases} \to \text{Yes} \Rightarrow + \\ \to \text{No} \Rightarrow - \end{cases}$$

inference

How do we find the params. w?

→ the learning problem

$$\hat{w} = \underset{w}{\operatorname{argmin}}\ L(\underbrace{X, Y}_{\text{set of training examples}}, w)$$

$$= \underset{w}{\operatorname{argmin}} \sum_{n=1}^{N} \text{loss}(x_n, y_n, w)$$

$$\text{loss}(x_n, y_n, w) \to -\log p(y_n | x_n; w) \to -y_n \log \hat{y}_n - (1-y_n)\log(1-\hat{y}_n)$$

Notation: let $\hat{y}_n = p(+|x_n; w)$

$p(-|x_n; w) = 1 - \hat{y}_n$

$$\hat{w} = \underset{w,b}{\arg\min} + \sum_{n=1}^{N} loss(x_n, y_n, w)$$

$$= \underset{w,b}{\arg\min} - \sum_{n=1}^{N} y_n \log(\hat{y}_n) + \cancel{y}(1-y_n)\log(1-\hat{y}_n)$$

$$\boxed{\hat{w} = \underset{w,b}{\arg\min} \left(- \sum_{n=1}^{N} y_n \log \underbrace{\sigma(w^T x_n + b)}_{\hat{y}_n} + (1-y_n) \log\big(1-\underbrace{\sigma(w^T x_n + b)}_{\hat{y}_n}\big) \right)}$$

Training a LR model

$\rightarrow Loss(w,b).$

$$\hat{w} = \underset{w,b}{\arg\min} \; Loss(w,b) \rightarrow \text{use GD.}$$