Classification

Regression $\rightarrow$ Continuous

$x \rightarrow y$

Classify $\quad x \rightarrow y \quad \{0, 1, 2, \cdots\}$

$x \rightarrow \{0, 1\}$

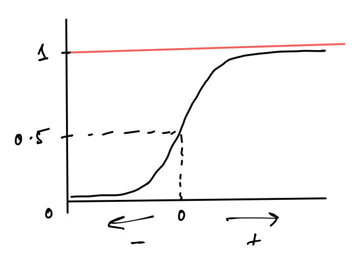
Probability

$P(y=1 \mid x;\theta)$

$h \rightarrow \underline{\theta^T x} = \sum_{i=0}^{d} \theta_i x_i$

Wrap

$g(\theta^T x) \rightarrow$ Sigmoid



$sig(\infty) \rightarrow 1$

$sig(-\infty) \rightarrow 0$

$sig(0) \rightarrow 0.5$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$y=1$ Positive
$y=0$ Neg.

$$P(y=1 \mid x;\theta) = h(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$P(y=0 \mid x;\theta) = \underline{1 - P(y=1 \mid x;\theta)}$$

$h(x) = g(\theta^T x)$

Loss function $= J(\theta) \downarrow \quad \uparrow$ performance

CS229 $\rightarrow$ MSE $\leftarrow$

Maximize likelihood of data.

Minimize negative likelihood of data.

Loss = negative $\underline{L}L$

$$L(y \mid x;\theta) = (h(x))^{y}(1-h(x))^{1-y}$$

$\quad \hookrightarrow$ One instance

$y=1 \leftarrow \boxed{h(x)}$

$y=0 \leftarrow \boxed{1-h(x)}$

$\underline{y=0} \quad L = \underbrace{(h(x))^{0}}(1-h(x))^{1-0} \leftarrow$

$= 1(1-h(x))$

$= \underline{1-h(x)\downarrow}$

$= \uparrow$

$y=1 \quad L = (h(x))^{1}(1-h(x))^{1-1}$

$= h(x)$

Total data likelihood

$$L(y \mid x;\underline{\theta}) = \prod_{i=1}^{m} \underline{(h(x^{(i)}))^{y^{(i)}}} \cdot \underline{(1-h(x^{(i)}))^{1-y^{(i)}}}$$

IID $\rightarrow$ Independent
Identically
distributed

$\log(ab) = \log a + \log b$

$$\log L = LL(\theta) = \sum_{i=1}^{m}\left\{\log(h(x^{(i)}))^{y^{(i)}} + \log(1-h(x^{(i)}))^{1-y^{(i)}}\right\}$$

$$= \sum_{i=1}^{m} \underline{y^{(i)}\log(h(x^{(i)})) + (1-y^{(i)})\log(1-h(x^{(i)}))}$$

max $\uparrow$ best θ

$-LL(\theta)\downarrow$ min best θ

Loss $= -LL(\theta)$

$J(\theta)$

$$\underbrace{\theta := \theta - \alpha \frac{\partial}{\partial \theta} J(\theta)}$$

$$\frac{\partial}{\partial \theta_i} LL = (y - h(x))\,x_i$$

Multiple classes

$y \in \{0, 1, \cdots, k\}$

$P(y=1 \mid x;\theta) = h(x)$

$P(y=0) = 1-h$

$g(\underline{\theta^T x})$
$\downarrow$ Rethink

$\hat{y} = [0, 1]$

$\sum p = 1$

Binary $\rightarrow$ Bernoulli $\rightarrow$ Coin Toss

$\phi = g(\theta^T x) \leftarrow$

Multiple $\rightarrow$ Multinomial $\rightarrow$ Dice $\{ \cdots k\}$

$\phi_0, \phi_1, \cdots, \phi_K$

$\theta_0^T x \leftarrow \theta_0$ is a vector

$\theta_1^T x \leftarrow \theta_1$

$\vdots$

$\theta_K^T x$

$$\text{softmax}\begin{bmatrix}\theta_0^T x \\ \theta_1^T x \\ \vdots \\ \theta_K^T x\end{bmatrix} = \begin{bmatrix}\dfrac{e^{t_0}}{\sum_{i=0}^{K} e^{t_i}} \\ \dfrac{e^{t_1}}{\sum_{i=0}^{K} e^{t_i}} \\ \vdots \\ \dfrac{e^{t_K}}{\sum_{i=0}^{K} e^{t_i}}\end{bmatrix}$$

$\rightarrow [0, 1]$

$[0, 1]$

$\leftarrow$ Sum $\downarrow$ 1

$t_0 = \theta_0^T x$

$$\phi_0 = g(\theta^T x) = \text{softmax}(\theta^T x) = \frac{e^{\theta_0^T x}}{\sum_{i=0}^{d} e^{\theta_i^T x}}$$

Loss to min

Score to max

$$L(y=K \mid x;\theta) = \phi_K = \frac{\exp(\theta_K^T x)}{\sum_{i=0}^{d}\exp(\theta_i^T x)}$$

One instance

$$LL(y=K \mid x;\theta) = \log(\exp \theta_K^T x) - \log\left(\sum_{i=0}^{d}\exp \theta_i^T x\right)$$

$$= \theta_K^T x - \log\left(\sum_{i=0}^{d}\exp \theta_i^T x\right)$$

$$-LL = \log\left(\sum_{i=0}^{d}\exp \theta_i^T x\right) - \theta_K^T x$$

$$\frac{\partial}{\partial \theta_K} - LL(\theta) = \frac{1}{\sum_{i=0}^{d}\exp(\theta_i^T x)} \cdot \sum_{i=0}^{d}\left(\frac{\partial}{\partial \theta_K}\exp(\theta_i^T x)\right) - x$$

$$= \frac{1}{\sum_{i=0}^{d}\exp(\theta_i^T x)} \sum_{i=0}^{d}\left(e^{\theta_i^T x}\underline{\frac{\partial}{\partial \theta_K}\theta_i^T x}\right) - x$$

$$= \frac{1}{\sum_{i=0}^{d}\exp(\theta_i^T x)} e^{\theta_K^T x} \cdot x - x$$

$$= \left\{\underline{\frac{e^{\theta_K^T x}}{\sum_{i=0}^{d}\exp(\theta_i^T x)}} - 1\right\} x$$

$$\frac{\partial}{\partial \theta_K} - LL(\theta) = (\phi_K - 1)\,x$$

$$\sum_{i=1}^{m}(\quad)$$

Accuracy $\quad$ 80 $\quad$ 20

F1-score

Recall —

Precision —

P $\rightarrow$ P, N

N $\rightarrow$ P, N

$\left\{\begin{array}{ll} P \rightarrow P & \text{True P} \\ P \rightarrow N & \text{False N} \\ N \rightarrow P & \text{False P} \\ N \rightarrow N & \text{True Negative} \end{array}\right.$

Recall $\rightarrow$ Out of all the positive instances, how many were predicted positive?

0 1 1 0 0 1 1

1 1 1 1 1 1 1 $\quad$ R = 1

Precision : In your predicted positives, how many are actually positive.

0 1 1 0 0 1 1

0 1 1 0 0 0 0 $\quad$ Precision = 1

$\underline{F1}$ = Balances Precision — Recall —

Covid-19

10 $\rightarrow$ 4
6 $\rightarrow$ 2

Recall $\uparrow$

Youtube Kids

4 $\rightarrow$ censor

8

3 $\quad$ 1 $\rightarrow$ Kid

Recall $\downarrow$

Precision $\uparrow$

---

1. Cannot linear regression
   - $\hookrightarrow$ Continuous
   - $\hookrightarrow$ Prefer Proba.
2. Binary
   - $\hookrightarrow$ $\theta^T x \rightarrow$ Prob. using sigmoid
3. How do we optimize
   - $\hookrightarrow$ Loss $\rightarrow$ Negative Log Likelihood
   - Score $\rightarrow$ Log Likelihood
4. Gradient descent.
5. Multiple classes.
   - $\hookrightarrow$ Multinomial $\rightarrow \phi_0\ \phi_1 \cdots \phi_K$
   - $\hookrightarrow$ $\phi_0 = \text{softmax}(\theta_0^T x)$
     θ_0 is a vector for class 0.
6. $-LL \rightarrow$ Loss
   GD
7. Metrics
   - $\hookrightarrow$ acc
   - $\hookrightarrow$ confusion matrix
   - $\hookrightarrow$ precision / recall
   - $\hookrightarrow$ ROC