Modeling binary outcomes

Deep Learning / Good fellow — Manning Publishers Chollet, Allaire R Deep Learning

pytorch — Tensorflow Google — Keras _ R keras

---

$\log\left\{\frac{P(y=1 \mid x_1, x_2 \cdots x_p)}{P(y=0 \mid x_1 \cdots x_p)}\right\} = \log \text{odds} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots \beta_p x_p = \eta$

$P(y=1 \mid x_1 \cdots x_p) = \frac{e^{\eta}}{1+e^{\eta}}$

$\log\{\text{odds}(y=1 \mid x_1 = x_{i+1}, \cdots x_p = x_p)\} = \beta_0 + \beta_1(x_i+1) + \cdots \beta_p x_p$

$\log\{\text{odds}(y=1 \mid x_1 = x_{i+1}, \cdots x_p = x_p)\} = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$

$1 - 2 = \beta_1$

$\beta_k$ = increase or decrease in the log odds of a success given a 1 unit change in $x_k$, holding the other regressors constant.

$e^{\beta_k}$ = the odds ratio for a 1 unit increase in $x_k$, holding the other regressors constant. $\frac{e^1}{e^2}$

$y$ = HTN — $x_k$ = Pack years — other regressors BMI, Anti HTN med, Age, Sex

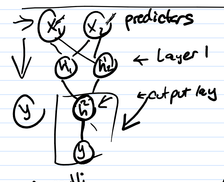
We expect a .1 increase in the log odds of HTN for every 1 year increase in pack years holding BMI, Anti HTN meds, Age, Sex constant

$e^{.1} \approx 1.1$ for the odds of hypertension are 10.5% higher for every pack years holding --- constant.

$e^{.1} \approx 1.1$ $\quad e^x \approx 1+x$ if $x \approx 0$

---

$P(y=1 \mid x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$



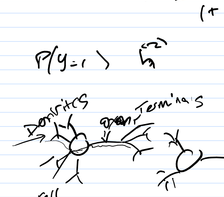

$h_1' = \frac{e^{\beta_0' + \beta_1' x_1 + \beta_2' x_2}}{1+e^{\beta_0' + \beta_1' x_1 + \beta_2' x_2}}$

$h_2' = \frac{e^{\beta_0'' + \beta_1'' x_1 + \beta_2'' x_2}}{1+e^{\beta_0'' + \beta_1'' x_1 + \beta_2'' x_2}}$

$h^{(1)} = \gamma_1 h_1 + \gamma_2 h_2$

Bernoulli (2), $\frac{e^{\gamma_1 h_1' + \gamma_2 h_2'}}{1+e^{\gamma_1 h_1' + \gamma_2 h_2'}}$

$P(y=1) > h_1^{(2)}$




---

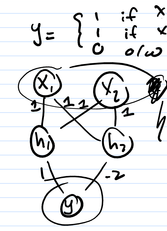
RELU = rectified linear unit

$g(z) = \max(0, z)$

$= \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$

$y = \text{xor}\{x_1, x_2\}$

$y = \begin{cases} 1 & \text{if } x_1 = 1 \text{ \& } x_2 = 0 \\ 1 & \text{if } x_1 = 0 \text{ \& } x_2 = 1 \\ 0 & \text{o/w} \end{cases}$




| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |

$h_1 = g(x_1 + x_2 + 0)$

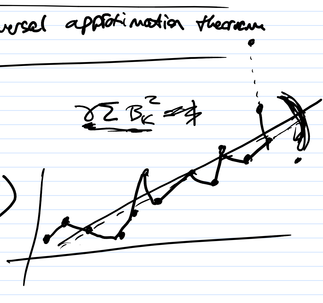
$h_2 = g(x_1 + x_2 - 1)$

$\hat{y} = h_1 - 2h_2$

| $x_1$ | $x_2$ | $h_1$ | $h_2$ | $h_1$ | $h_2$ | $\hat{y}$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | -1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 2 | 1 | 2 | 1 | 0 |

---

Universal approximation theorem




---

Regularization

Random dropout

(approximates a bootstrap procedure)