29/04/2020

$$\mathrm{Ber}(y_n \mid \sigma(w^T x_n)) = \sigma(w^T x_n)^{y_n} (1 - \sigma(w^T x_n))^{1-y_n}$$

$$\log(\mathrm{Ber}(y_n \mid \sigma(w^T x))) = y_n \log(\sigma(w^T x_n)) + (1-y_n) \log(1-\sigma(w^T x_n))$$

$$1 - \frac{1}{1+e^{-w^T x}} = \frac{1+e^{-w^T x} - 1}{1+e^{-w^T x}} = \frac{1}{1+e^{w^T x}}$$

3 | MULTIPLE CHOICE ANONYMOUS | Edit

The following table lists a dataset of the scores students achieved on an exam described in terms of whether the student studied for the exam (Studied) and the energy level of the lecturer when grading the student's exam (Energy)

| ID | Studied | Energy | Score |
|---|---|---|---|
| 1 | yes | tired | 65 |
| 2 | no | alert | 20 |
| 3 | yes | alert | 90 |
| 4 | yes | tired | 70 |
| 5 | no | tired | 40 |
| 6 | yes | alert | 85 |
| 7 | no | tired | 35 |

Which of the two descriptive features should we use as the testing criterion at the root node of a decision tree to predict students' scores?

**A.** Studied

**B.** Energy

```
In [79]: studied_no = np.array([20, 40, 35])

In [80]: studied_yes = np.array([65, 90, 70, 85])

In [81]: energy_tired = np.array([65, 70, 40, 35])

In [82]: energy_alert = np.array([20, 90, 85])

In [83]: (4/7)*np.var(studied_yes, ddof=1) + (3/7)*np.var(studied_no, ddof=1)
Out[83]: 127.38095238095238

In [84]: (4/7)*np.var(energy_tired, ddof=1) + (3/7)*np.var(energy_alert, ddof=1)
Out[84]: 829.7619047619047
```