16.2) Cross-Entropy, Sigmoid, and Softmax

Vitor Kamada

August 2019

Entropy: Measure (Im)purity

$$S = [9+, 5-]$$

$$Entropy(S) = -p_{\oplus} log_2 p_{\oplus} - p_{\ominus} log_2 p_{\ominus}$$

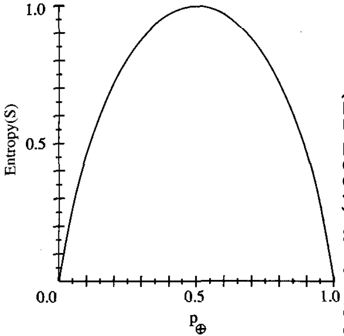
$$= -\tfrac{9}{14} log_2(\tfrac{9}{14}) - \tfrac{5}{14} log_2(\tfrac{5}{14}) = 0.94$$

$$Z = [14+, 0-]$$

$$Entropy(Z) = -\tfrac{14}{14} log_2(\tfrac{14}{14}) - \tfrac{0}{14} log_2(\tfrac{0}{14})$$

$$= 0$$

Entropy $\in$ [0,1]



Mitchell (1997:57)

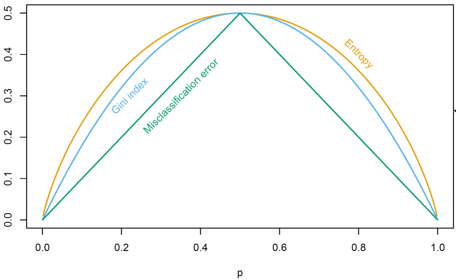
$$= -0.5log_2(0.5) - 0.5log_2(0.5)$$

$$= -0.5(-1) - 0.5(-1) = 1$$

Cross-Entropy (Deviance)$= -\sum_{k=1}^{K} \hat{p}_k log(\hat{p}_k)$

Correct Classified: $\hat{p}_k = \frac{1}{N} \sum_{n=1}^{N} I(y_i = k)$

Misclassification Error: $1 - \hat{p}_k$



Hastie et al. (2017:309)

Gini Index:$- \sum_{k=1}^{K} \hat{p}_k(1 - \hat{p}_k)$

Cross-Entropy (D) is more Sensitive to Changes in the Node Probabilities than Classification Error (E)

a)

$$\begin{array}{ccc} & (400, 400) & \\ \swarrow & & \searrow \\ (300, 100) & & (300, 100) \end{array}$$

$$E_a = (\tfrac{1}{2})(\tfrac{100}{400}) + (\tfrac{1}{2})(\tfrac{100}{400}) \\ = 0.25$$

---

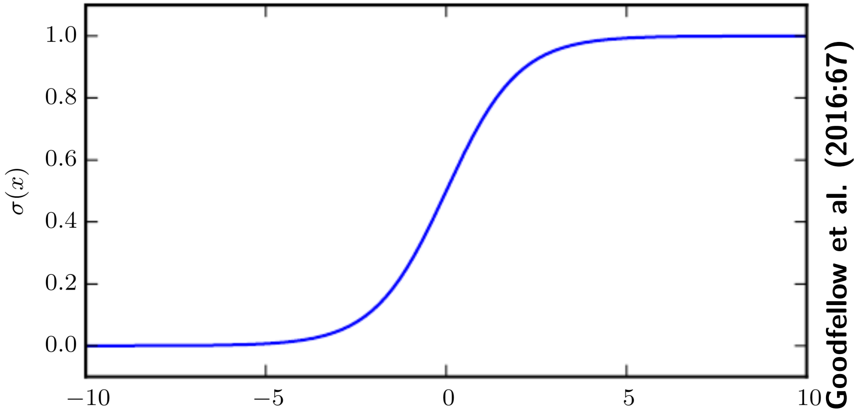
b)

$$\begin{array}{ccc} & (400, 400) & \\ \swarrow & & \searrow \\ (200, 400) & & (200, 0) \end{array}$$

$$E_b = (\tfrac{3}{4})(\tfrac{200}{600}) + (\tfrac{1}{4})(\tfrac{0}{200}) \\ = 0.25$$

$$D_a = E_a \text{ but } D_b < E_a$$

Logistic Sigmoid



$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$1 - \sigma(x) = \sigma(-x)$$

Properties of Logistic Sigmoid

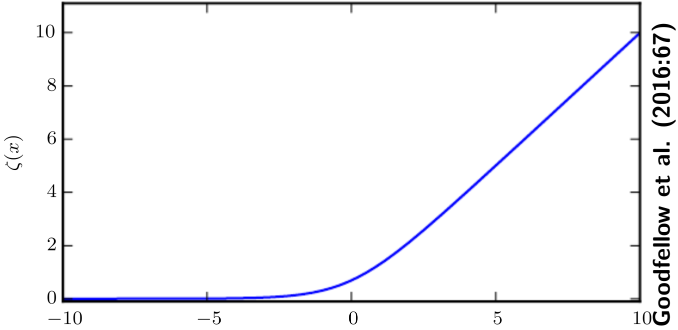
$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

$$1 - \sigma(x) = \frac{e^x+1}{e^x+1} - \frac{e^x}{e^x+1} = \frac{1}{e^x+1}$$

$$\frac{d}{dx}\sigma(x) = \frac{e^x(e^x+1)-e^x(e^x)}{(e^x+1)^2} = \frac{e^x}{(e^x+1)^2}$$

$$= \sigma(x)[1 - \sigma(x)]$$

Softplus Function

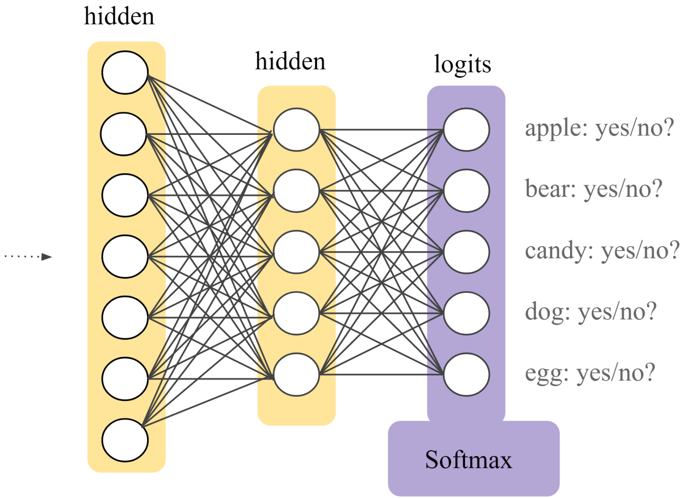


$$\zeta(x) = log(1 + e^x)$$

$$x^+ = max(0, x)$$

$$\frac{d\zeta(x)}{dx} = \frac{e^x}{e^x+1} = \sigma(x)$$

Softmax Layer



Source: https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax

Softmax Function (σ)

$$\mathbf{z} = \mathbf{W}^{\mathbf{T}}\mathbf{h} + \mathbf{b}$$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \textbf{ for } i = 1, \ldots K$$

$$log[\sigma(z)_i] = z_i - log \sum_{j=1}^{K} e^{z_j}$$