Baseline Models Difference

As a starting point in Multinomial Logistic Regression we can choose any class to be the **baseline**

Example

- Say we have A, B, C classes
- We pick B to be the baseline
  The model estimate

$$\log\left(\frac{\Pr(\mathrm{Y=A})}{\Pr(Y=B)}\right) = \vec{X}\vec{\beta}_A$$

$$\log\left(\frac{\Pr(\mathrm{Y=C})}{\Pr(Y=B)}\right) = \vec{X}\vec{\beta}_C$$

After adding exponent for both sides we get

$$\frac{P(Y=A)}{P(Y=B)} = e^{\vec{X}\vec{\beta}_A}$$

$$\frac{P(Y=C)}{P(Y=B)} = e^{\vec{X}\vec{\beta}_C}$$

Let $S_A = e^{\vec{X}\vec{\beta}_A} S_C = e^{\vec{X}\vec{\beta}_C}$

$$P(A) + P(B) + P(C) = 1$$

$$P(C) = S_C P(B)$$

$$P(A) = S_A P(B)$$

$$P(B) = P(B)$$

$$S_A P(B) + P(B) + S_C P(B) = 1$$

$$P(B)(S_A + S_C + 1) = 1$$

$$P(B) = \frac{1}{S_A + S_C + 1} = \frac{1}{1 + e^{\vec{X}\vec{\beta}_A} + e^{\vec{X}\vec{\beta}_C}}$$

$$P(A) = \frac{e^{\vec{X}\vec{\beta}_A}}{1 + e^{\vec{X}\vec{\beta}_A} + e^{\vec{X}\vec{\beta}_C}}$$

$$P(C) = \frac{e^{\vec{X}\vec{\beta}_C}}{1 + e^{\vec{X}\vec{\beta}_A} + e^{\vec{X}\vec{\beta}_C}}$$

Why Baseline is needed?

- Cause probabilities must sum to 1

$$P(Y=A) + P(Y=B) + P(Y=C) = 1$$

- Only $K-1$ sets of classes are needed to model K model