Classification: Predict y from x

for $y \in \{1, 2, \ldots, K\}$.

$$\mathbb{P}(Y=g \mid X=x) \propto \mathbb{P}(X=x \mid Y=g)\, \pi_g$$

Suffices to find $(K-1)$ groups

If we look to model the group assignments using a Logistic Regression.

$$\log \frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=K \mid X=x)} = x^T \beta_k$$

$$\Longleftrightarrow$$

$$\mathbb{P}(Y=k \mid X=x) = \frac{\exp(\beta_k^T x)}{1 + \sum_{l=1}^{K-1} \exp(\beta_l^T x)}$$

$$\mathbb{P}(Y=k \mid X=x) = \frac{1}{1+\sum_{\ell=1}^{K-1} \exp(\beta_\ell^T x)}$$

To emphasize that $\mathbb{P}(Y=k \mid X=x)$ depends on all $\beta_\ell^T x$ we write $\theta = \{\beta_\ell^T x : 1 \leq \ell \leq K-1\}$.

$$\mathbb{P}(G=k \mid X=x) = p_k(x; \theta)$$

We'll fit this using IRLS and the update

$$\beta^{(new)} = \beta^{(old)} - \left(\frac{\partial^2 \ell(\beta)}{\partial\beta\,\partial\beta^T}\right)^{-1} \frac{\partial \ell(\beta)}{\partial\beta}$$

for $\ell(\beta) = \log \mathcal{L}(\beta; \mathcal{T})$.

$$= \sum_{i=1}^{n} \left\{ y_i \cdot \beta^T x_i - \log\left(1 + e^{\beta^T x_i}\right) \right\}$$

Use Newton Raphson to minimize this logistic regression IRLS.