Logistic Regression via Gradient Descent

Minimize logloss by setting the gradient of the loss function equal to zero and solving:

$$\nabla_\beta L(\beta) = \nabla_\beta -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log p_i + (1-y_i)\log(1-p_i)\right]$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \nabla_\beta \log p_i + (1-y_i)\nabla_\beta \log(1-p_i)\right]$$

Now, let $\phi(z) = \frac{1}{1+e^{-z}} = \text{logistic}(z)$

Then $\frac{d}{dz}\phi(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \phi(z)(1-\phi(z))$

$$\nabla_\beta p_i = \nabla_\beta\left(\frac{1}{1+e^{-x_i^T\beta}}\right) = \nabla_\beta\, \phi(x_i^T\beta)$$

$$= \phi(x_i^T\beta)(1-\phi(x_i^T\beta))\,\vec{x}_i \quad \text{by Chain Rule}$$

$$= p_i(1-p_i)\,\vec{x}_i$$

Hence

$$\nabla_\beta L(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \nabla_\beta \log p_i + (1-y_i)\nabla_\beta \log(1-p_i)\right]$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \frac{\nabla_\beta p_i}{p_i} - (1-y_i)\frac{\nabla_\beta (1-p_i)}{1-p_i}\right]$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i (1-p_i) x_i - (1-y_i) p_i x_i\right]$$

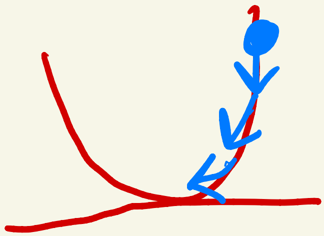
$$= -\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)\, x_i$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \phi(x_i^T\beta)\, x_i\right)$$

$\blacksquare$

* Setting $\nabla_\beta L(\beta) = 0$ has no known closed form solution.

So, use <u>Newton Raphson</u> or <u>Gradient descent</u> to approximate $\hat{\beta} = \underset{\beta}{\arg\min}\, L(\beta)$.



Gradient Descent

Iterate until convergence of $\vec{\beta}$,

i.e. until $\|\vec{\beta}^{(t)} - \vec{\beta}^{(t+1)}\| < \varepsilon$:

$$\boxed{\vec{\beta}^{(t+1)} = \vec{\beta}^{(t)} + K \cdot \sum_{i=1}^{n} \left(y_i - \phi(x_i^T \beta)\right) \vec{x}_i}$$

K = learning Rate

Anyone Recognize K ??

ELO

* $$ELO^{(t+1)} = ELO^{(t)} + K\left(\mathbb{1}(\text{win}) - \mathbb{P}(\text{win})\right)$$

One iteration of gradient descent in logistic Regression
for Bradley Terry Power scores
is one ELO update!