Lecture September 24

Linear regression

$$y_i = f(x_i) + \varepsilon_i$$ ← Gaussian PDF

Logistic regression

$$y_i = p(x_i) + \tilde{\varepsilon}_i$$ ← Binomial distribution (Binary case)

$$p(x_i) = p(y_i | x_i \beta)$$

$$= \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = p_i$$

$$y_i = 1$$

$$y_i = 0 \quad ; \quad p(y_i = 0) = 1 - p(y_i = 1)$$

MLE

$$P(y | X\beta) = \prod_{i=0}^{n-1} p_i^{y_i} (1 - p_i)^{1 - y_i}$$

$$\text{cost } C(\beta) = - \sum \left[y_i \log p_i + (1 - y_i) \log (1 - p_i) \right]$$

$$\frac{\partial C}{\partial \beta_0} = - \sum_{i=0}^{n-1} (y_i - p_i) = \sum_i (p_i - y_i)$$

$$\frac{\partial C}{\partial \beta_1} = \sum x_i (p_i - y_i)$$

$$\frac{\partial C}{\partial \beta} = X^T(p-y) = \underline{g(\beta)}$$

$$\frac{\partial^2 C}{\partial \beta \partial \beta^T} = \underset{\text{Hessian matrix}}{\underline{H}} = \sum_i p_i(1-p_i)\, x_i x_i^T$$

$$= X^T W X$$

$$W = \operatorname{diag}\left(p_i(1-p_i)\right)$$

$$\frac{\partial^2 C}{\partial \beta_0 \partial \beta_1} \qquad \frac{\partial^2 C}{\partial \beta_0^2} \qquad \frac{\partial^2 C}{\partial \beta_1^2}$$

$$\frac{\partial C}{\partial \beta} = 0 = X^T(p-y)$$

p depends in a non-linear way on β.

classification measures

TP = True positive, equ with a correct hit

TN = True negative

FP = False positive, false alarm

FN = False negative

Total set contains -n- entries.

accuracy score

$$= \frac{\sum_{i=0}^{n-1} I(y_i = \tilde{y}_i)}{n}$$

$I(y_i = \tilde{y}_i) = 1$, zero else

$$= \frac{\sum TP + \sum TN}{n}$$

precision $\frac{TP}{TP + FP}$ sum over all

Recall : TP

$$\frac{\quad}{TP + FN}$$

$$F1\text{-score} = \frac{2\,TP}{TP + \frac{1}{2}(FP + FN)}$$

Confusion matrix

TP	FP
FN	TN

ROC - curves

True positive rate : $\frac{TP}{TP + FN}$

is plotted against

False positive rate $= \frac{FP}{FP + FN}$

Gains curve

$$\frac{\text{count } TP + \text{count } FP}{\text{count of all observations}}$$

$= x\text{-axis}$

y-axis : $\dfrac{\text{count TP}}{\text{count TP} + \text{count FN}}$

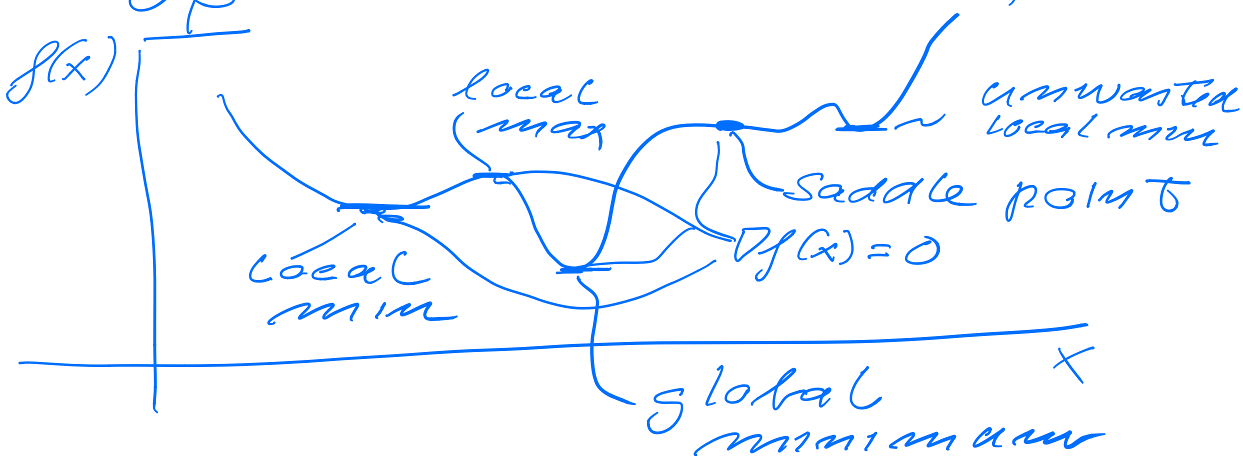
TP = count TP

optimization/minimization/maximization

$$\hat{\beta}^{OLS} = \hat{\beta}^{OLS}_{opt} = (X^T X)^{-1} X^T y$$

$$\hat{\beta}^{opt}_{LogReg} = ?$$

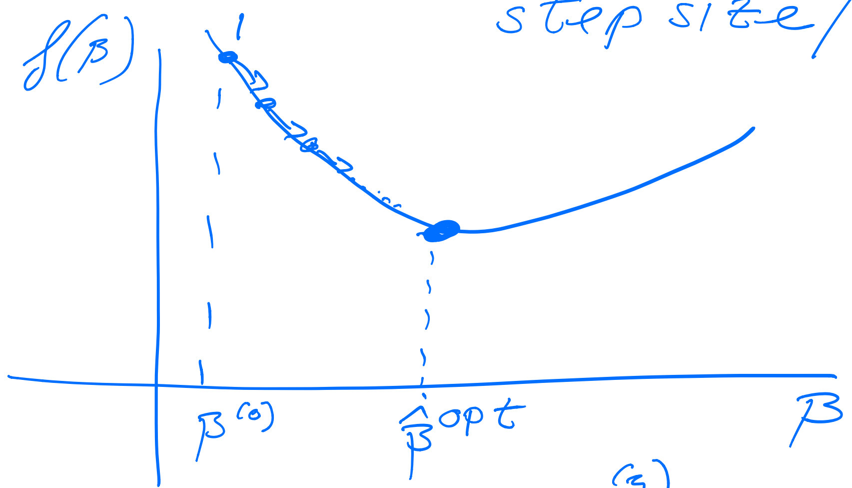
$$\frac{\partial C}{\partial \beta} = 0 = X^T (p - y)$$



Iterative process:

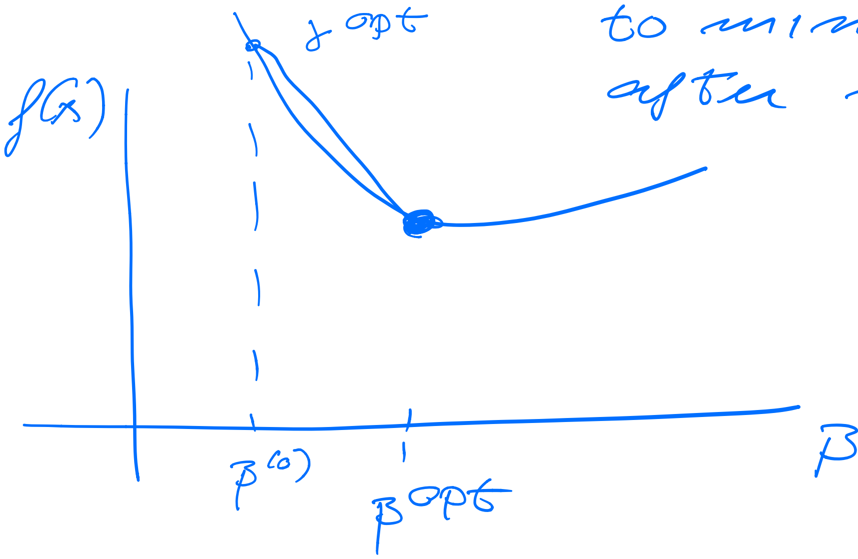
$$\beta^{(n+1)} = \beta^{(n)} - \gamma^{(n)} g(\beta^{(n)})$$

step size / learning rate,

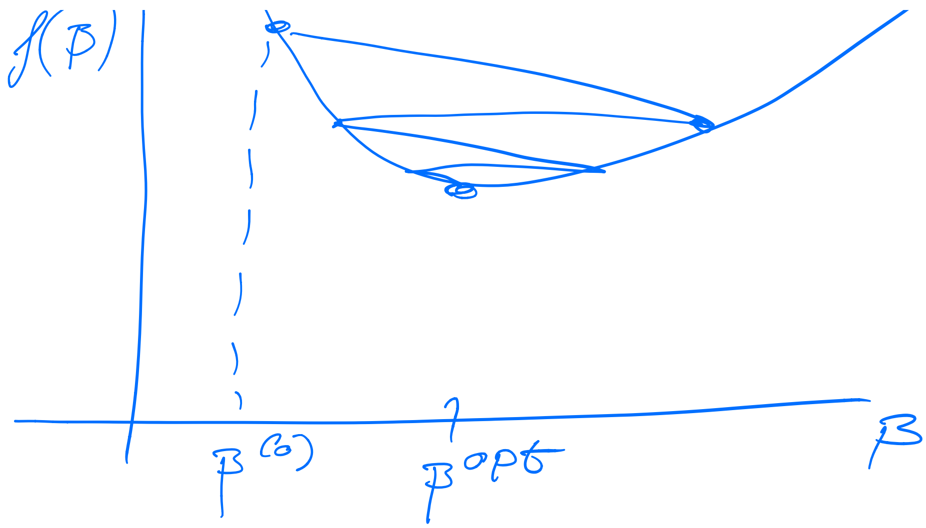


— $\gamma^{(n)}_{opt}$

$\gamma < \gamma^{opt}$, may not converge to minimum after n-iteration



if $\gamma^{opt} < \gamma < 2\gamma^{opt}$
the search will oscillate between both sides till we reach a minimum



$\gamma > 2\gamma^{opt}$, no solution.

Eigenvalues of Hessian matrix can be used to show

$$\gamma < \frac{2}{\lambda_{max}}$$

condition number max $\left| \frac{\lambda_{max}}{\lambda_{min}} \right|$

Newton-Raphson root searching method.

$f(s) = 0$ $\quad (g(\beta))$

$$= f(x) + (s-x) f'(x) + \frac{(s-x)^2}{2!} f''(x) + O((s-x)^3)$$

Skip higher-order (2nd and on) terms

$$f(x) + (s-x) f'(x) = 0 \Rightarrow$$

$$s = x - f(x)/f'(x)$$

Lend itself to iterations

$$x^{(n+1)} = x^{(n)} - f(x^{(n)}) / f'(x^{(n)})$$

two-variables

x_1, x_2

$$f_1(x_1, x_2) = 0$$

$$f_2(x_1, x_2) = 0$$

$$\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

$$\left[\begin{array}{l} \dfrac{\partial C}{\partial \beta_0} = \sum_i (p_i - y_i) = \underline{g_0}(\beta_0, \beta_1) = 0 \\ \dfrac{\partial C}{\partial \beta_1} = \sum_i x_i (p_i - y_i) = \underline{g_1}(\beta_0, \beta_1) \end{array}\right.$$

Taylor expand

$$f_1(x_1 + h_1, x_2 + h_2) = f_1(x_1, x_2) + h_1 \partial f_1/\partial x_1 + h_2 \partial f_1/\partial x_2 + \ldots$$

$$f_2(x_1+h_1, x_2+h_2) = f_2(x_1, x_2) + h_1\, \partial f_2/\partial x_1 + h_2 \frac{\partial f_2}{\partial x_2} + \dots$$

Define Jacobian matrix

$$J = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix}$$

iterative scheme

$$\begin{bmatrix} x_1^{(n+1)} \\ x_2^{(n+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \end{bmatrix} - J^{-1} \begin{bmatrix} f_1(x_1^{(n)}, x_2^{(n)}) \\ f_2(x_1^{(n)}, x_2^{(n)}) \end{bmatrix}$$

$$\frac{f(x^{(n)})}{f'(x^{(n)})}$$

$\beta_0 \quad \beta_1$

$$\underline{J} = H = \begin{bmatrix} \partial g_0/\partial \beta_0 & \partial g_0/\partial \beta_1 \\ \partial g_1/\partial \beta_0 & \partial g_1/\partial \beta_1 \end{bmatrix}$$

$(n+1) \qquad (n)$

$$\begin{bmatrix} \beta_0^{(n+1)} \\ \beta_1^{(n+1)} \end{bmatrix} = \begin{bmatrix} \beta_0^{(n)} \\ \beta_1^{(n)} \end{bmatrix} - H^{-1} \times \begin{bmatrix} g_0(\beta_0^{(n)}, \beta_1^{(n)}) \\ g_1(\beta_0^{(n)}, \beta_1^{(n)}) \end{bmatrix}$$

$$\boxed{\beta^{(n+1)} = \beta^{(n)} - \left(\underline{H}^{-1} g\right)\Big|_{\beta = \beta^{(n)}}}$$

$$\boxed{\beta^{(n+1)} = \beta^{(n)} - \gamma^{(n)} g(\beta^{(n)})}$$

$$\sum_{i=0}^{n-1} (p_i - y_i) x_i$$

ML bottlenecks.

Taylor-expand $c(\beta)$

$c(\beta) \quad c(\beta^{(n)})$

$$C(\beta) = \underline{C(\beta^{(n)})} +$$

$$g^{(n)T}(\beta - \beta^{(n)}) + \frac{1}{2}(\beta - \beta^{(n)})^T$$

$$\underline{H}^{(n)}(\beta - \beta^{(n)})$$

$$= \underline{\alpha + \beta^T A \beta + \delta^T \beta}$$

$$A = \frac{1}{2} H^{(n)}$$

$$\delta = g^{(n)} - H^{(n)} \beta^{(n)}$$

$$\alpha = C(\beta^{(n)}) - g^{(n)T} \beta^{(n)}$$

$$+ \frac{1}{2} \beta^{(n)T} H^{(n)} \beta^{(n)}$$

$$\underline{\beta^{opt} = -\frac{1}{2} A^{-1} \cdot \delta}$$