Lecture November 22

Confusion matrix

TP 0.98	FP 0.02
FN 0.07	TN 0.93

TP = TRUE POSITIVE

TN = TRUE NEGATIVE CORRECT REJECTION

FP = False positive, false alarm

FN = False negative

accuracy score =

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

TRUE POSITIVE RATE = TPR

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

FALSE positive rate FPR

$$= \frac{FP}{}$$

FP+TN

Gradient Methods

Logistic regression

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

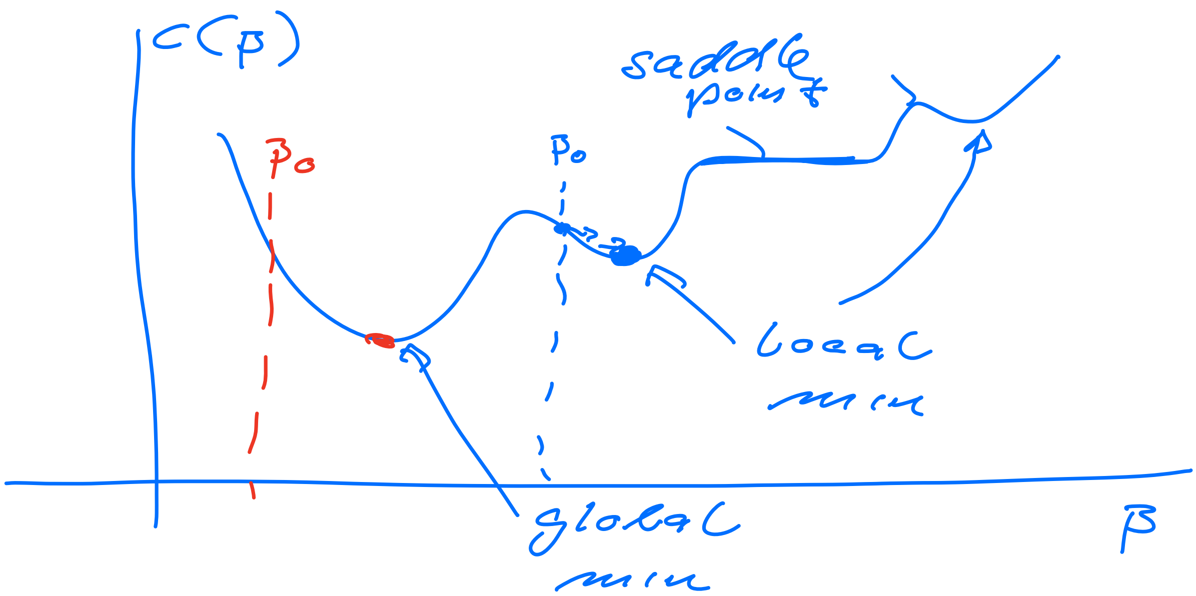
$p \in \mathbb{R}^n \quad y \in \mathbb{R}^n$

$X \in \mathbb{R}^{n \times p}$

$$\frac{\partial^2 C}{\partial \beta \partial \beta^T} = H = X^T W X = H(\beta)$$

$$w_{ii} = p_i(1-p_i)$$

(linear reg $H \propto X^T X$)



optimal $\hat{\beta} = \beta_{k+1} = \beta_k - H^{-1}(\beta_k)\, g(\beta_k)$

$\beta_k = \beta$

$|\beta_{k+1} - \beta_k| \le \varepsilon \approx 10^{-10}$

Taylor expand

$$c(\hat{\beta}) \cong c(\beta) + (\hat{\beta}-\beta)^T g(\beta) + \frac{1}{2}(\hat{\beta}-\beta)^T H(\beta)(\hat{\beta}-\beta)$$

$b = \hat{\beta} - \beta$

$$c(\hat{\beta}) = c(\beta) + b^T g + \frac{1}{2} b^T H b$$

$$\frac{\partial c}{\partial b'} = 0 = Hb + g \Rightarrow$$

$$b = H^{-1} g \Rightarrow$$

$$\hat{\beta} - \beta = H^{-1} g \Rightarrow$$

$$\hat{\beta} = \beta - H^{-1}(\beta)\, g(\beta)$$

Newton's method is a minimization of a func

$$f(x) = \frac{1}{2} x^T A x + x^T b + c$$

$$\frac{\partial f}{\partial x^T} = Ax + b = 0 \quad \Rightarrow$$

$$Ax = -b$$

$$\hat{\beta} = \beta - H^{-1} g \simeq \beta - \eta g(\beta)$$

$$(\beta_{k+1} = \beta_k - \eta_k g(\beta_k))$$

$$C(\hat{\beta}) = C(\beta) - \eta g^T g$$
$$+ \frac{1}{2} \eta^2 g^T H g$$

one-dim

$$C(\hat{\beta}) = C(\beta) - \eta g^2 + \frac{1}{2} \eta^2 g^2 H$$

$C(\beta)$ = original/start value

ηg^2 = improvement due to the slope

$\frac{1}{2} \eta^2 g^2 H$ = correction due to curvature.

$$\frac{\partial C}{\partial \eta} = 0 = -g^T g + \eta g^T H g$$

$$\Rightarrow \quad \eta_k = \frac{g^T g(\beta_k)}{g^T H g(\beta_k)}$$

suppose $Hg = \lambda g$

smallest $\eta = \dfrac{1}{\lambda_{max}}$

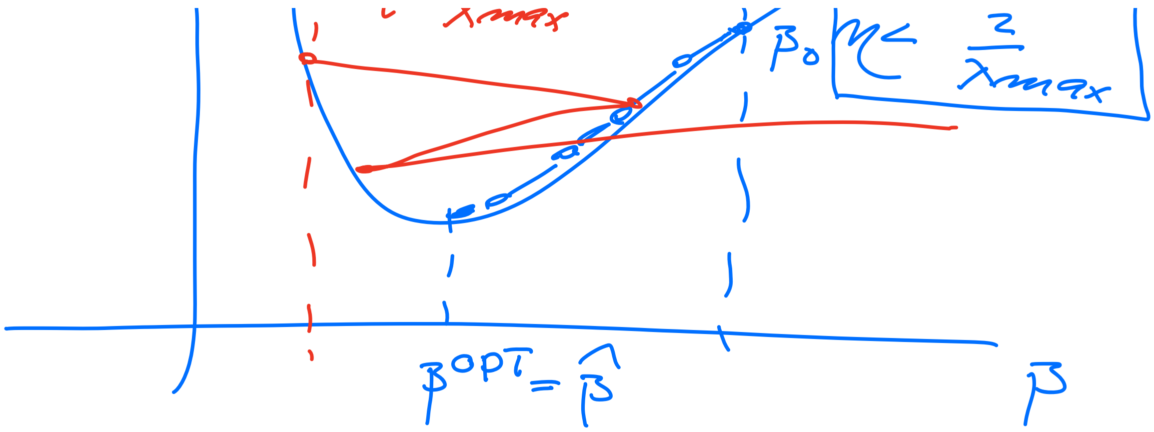
largest $\eta = \dfrac{1}{\lambda_{min}}$

can show, to achieve convergence

$\eta < \dfrac{2}{\lambda_{max}}$ $\quad$ λ_{max} is largest eigenvalue of H.

The standard gradient descent

$$\beta_{k+1} = \beta_k - \eta_k \nabla_\beta C(\beta_k)$$



$$f'' \approx \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

$$f' \approx \frac{f(x+h) - f(x-h)}{2h}$$