Lecture September 30

Classification notation

accuracy score

$$n = \sum_{i=0}^{n-1} (\text{total population})_i \qquad y_i = 0 \vee y_i = 1$$

$$= \sum TP + \sum TN$$

TP = True positive, eqv with hit

TN = True Negativ, eqv with correct rejection

FP = False positive, false alarm

FN = False negativ, eqv with miss

Confusion Matrix

TP	FP
FN	TN

accuracy score:

$$\frac{\sum \text{correctly classified}}{n}$$

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

Additional quantity

<u>True positive rate</u>:

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

False positive rate:

$$\frac{FP}{FP+TN} = FPR = 1 - TNR$$

TNR = True negative rate

$$= \frac{TN}{TN+FP}$$

GAINS CURVE :

x-axis : $\dfrac{\text{count } TP + \text{count } FP}{n}$

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

ROC - curve

True positive rate against False positive rate

Gradient Methods

* Linear Regression

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

$$\frac{\partial C(\beta)}{\partial \beta} = -\frac{2}{n} X^T (X\beta - y) = 0$$

$$\frac{\partial^2 C(\beta)}{\partial \beta \partial \beta^T} = \frac{2}{n} \underbrace{X^T X}_{\text{Hessian}}$$

$$(X^T X)_{ij} = \frac{\partial^2 C}{\partial \beta_i \partial \beta_j^T}$$

Logistic regression

$$p_i = p(y_i = 1\, x_i | \beta)$$

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

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

$W_{ii} = (1-p_i)\, p_i \qquad W_{ij} = 0$

$ij \quad i \neq j$

For finding optimal β-values, we solve iteratively (Newton-Raphson)

$$\beta^{(n+1)} = \beta^{(n)} - \left[H^{-1} \frac{\partial C}{\partial \beta} \right]_{\beta = \beta^{(n)}}$$

$$\left| \beta^{(n+1)} - \beta^{(n)} \right| < \delta \sim 10^{-10}$$

Gradient descent

$H^{-1} \longrightarrow \eta$ = learning rate

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

$$\frac{\partial C}{\partial \beta}\Big|_{\beta=\beta^{(n)}}$$

Digression: Linear regression

$$H = X^T X$$

$$X = U \Sigma V^T$$

$$(X^T X) V = \Sigma^2 V$$

$$\sigma_0^2 > \sigma_1^2 > \dots \sigma_{p-1}^2 > 0$$

condition number of a matrix $\max\limits_{i,j} \left| \frac{\lambda_i}{\lambda_j} \right|$

Taylor expand:

$$C(\hat{\beta}) \simeq C(\beta^{(n)}) +$$

$$\left(\hat{\beta} = \beta^{(n+1)} = \beta^{(n)} - \underbrace{\eta g^{(n)}} \right)$$

$$\left(\hat{\beta} - \beta^{(n)} \right) g^{(n)} +$$

$$\frac{1}{2} \left(\hat{\beta} - \beta^{(n)} \right)^T H \left(\hat{\beta} - \beta^{(n)} \right)$$

(1- Dim ;

$$c(\hat{\beta}) = c(\beta^{(n)}) + (\hat{\beta} - \beta^{(n)}) g^{(n)}$$

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

$$H = \left. \frac{\partial^2 c}{\partial \beta \, \partial \beta^T} \right|_{\beta = \beta^{(n)}}$$

Recipe ;

$$\hat{\beta} = \beta^{(n)} - \eta \underbrace{\frac{\partial c}{\partial \beta}}_{g^{(n)}} \Big|_{\beta = \beta^{(n)}}$$

$$c(\beta^{(n)} - \eta g^{(n)}) \simeq c(\beta^{(n)})$$

$$- \eta (g^{(n)})^T g^{(n)}$$

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

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

$$\hat{\eta} = \frac{g^T g}{g^T H g}$$

g is an eigenvector of H $\qquad Hg = \lambda g$

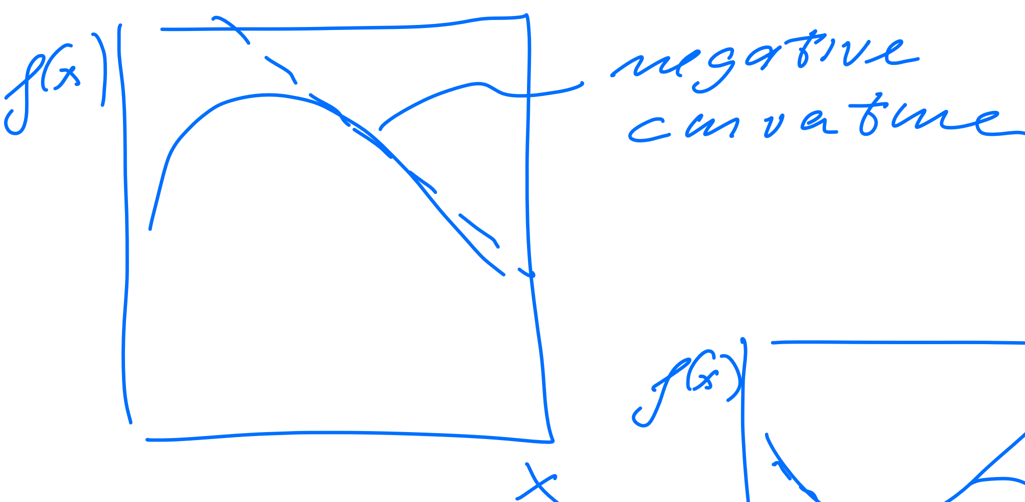
$$\hat{\eta} = \frac{1}{\lambda}$$

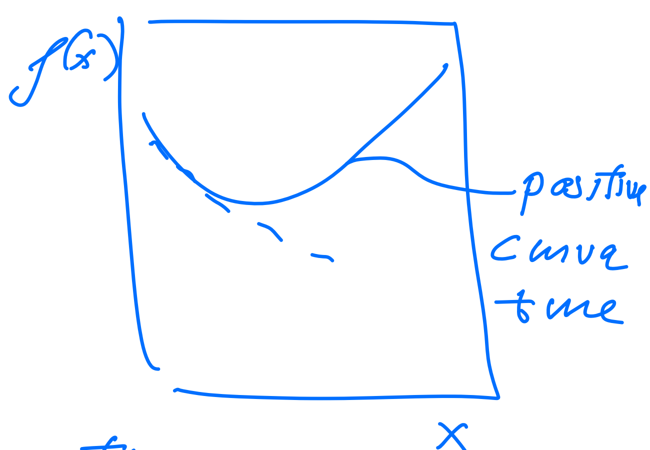
Three Terms:

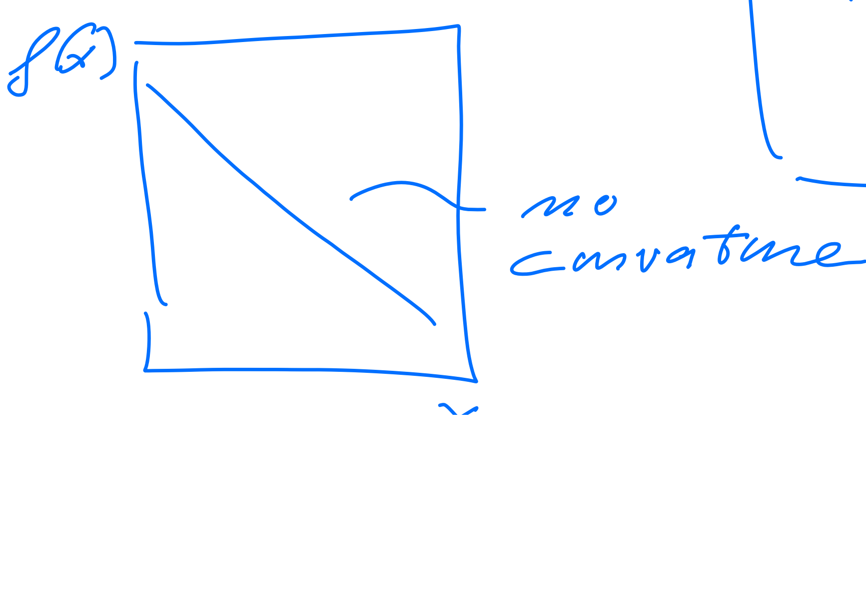
$-C(\beta^{(n)})$ = const value

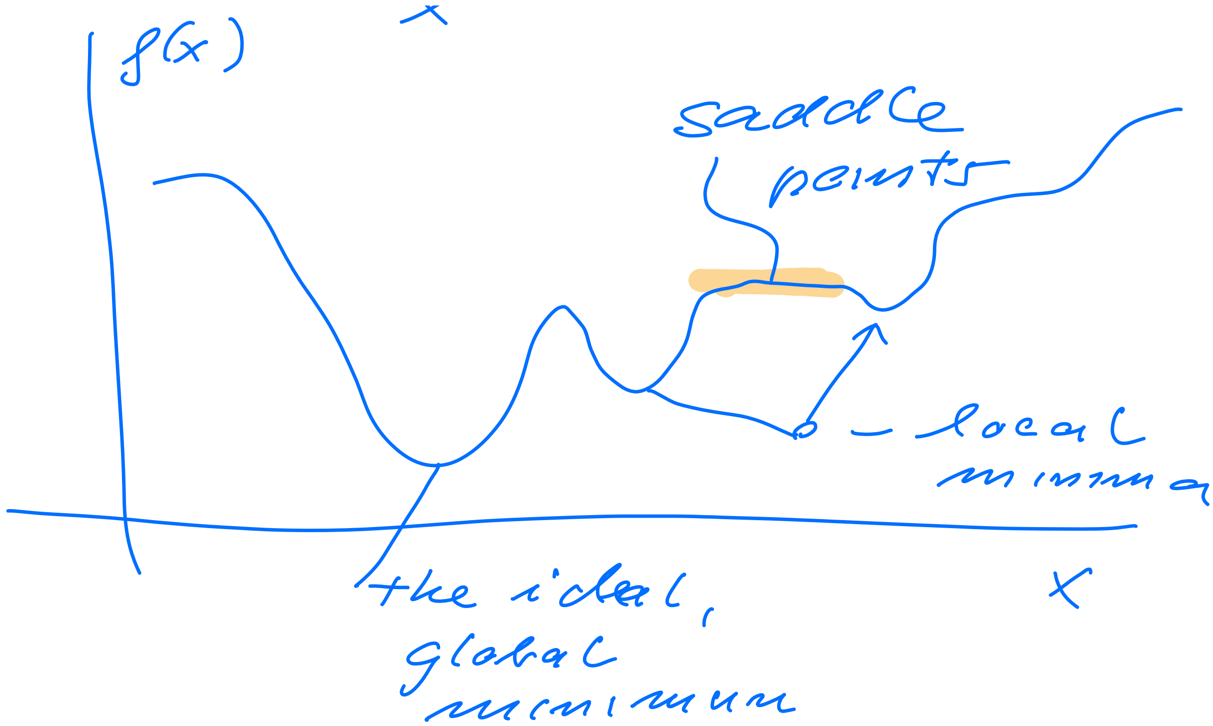
$-\eta g^T g$ = improvement to the slope of the function $C(\beta)$

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









$f'(x) = 0 \wedge f''(x) > 0$ local min

$f'(x) = 0 \wedge f''(x) < 0$ local max

$f''(x) = 0$, inconclusive

convex optimization

- Linear Regression
- Logistic — " —
- support vector machines