Erasmus+ course on machine learning, November 12, 2023

Taylor expanded $C(\beta)$ as function of $\hat{\beta} - \beta^{(m)}$

$$\nabla C(\beta) = 0$$

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

$$H(\beta) = \frac{\partial^2 C}{\partial \beta \partial \beta^T} \qquad g(\beta) = \nabla C(\beta)$$

$$\left(f(x) = x^T b - \frac{1}{2} x^T A x \right)$$

$$\nabla f(x) = 0 \Rightarrow Ax = b$$

Logistic regression

$$g(\beta) = X^T (p - y) \quad H = X^T W X$$

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

Expand around

$$\beta^{(n)} - \gamma g^{(n)} \quad | \; g^{(n)} = g(\beta^{(n)})$$

↑ replaces $+1$

γ = learning rate

$$C(\beta^{(n)} - \gamma g^{(n)}) \approx$$

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

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

↑ 2nd derivative

Take derivative wrt γ; $g^{(n)} \to g$

$H^{(n)} \Rightarrow H$

$$\gamma = \frac{g^T g}{g^T H g}$$ (parallel to steepest descent)

$$\beta^{(n)} - \frac{g^{T(n)} g^{(n)}}{g^{T(n)} H^{(n)} g^{(n)}} g^{(n)}$$

γ

assume

$$Hg = \lambda g$$

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

The convergence criterion is $\gamma < \frac{2}{\lambda_{max}}$

λ_{max} is largest eigenvalue of H

Gradient descent

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

ordinary least squares

$$g^{(m)} = X^T(X\beta - y)$$

algorithm of ordinary least squares

- initialize randomly $\beta^{(0)}$

$= [\beta_0^{(0)}, \beta_1^{(0)}, \ldots \beta_{p-1}^{(0)}]$

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

- - - $y \in \mathbb{R}^n$

specify iteration stop criterion -

define $\gamma = 0.01$

for $i = 1$, iter

calculate g

Do update

$\beta \leftarrow \beta - \gamma g$

stop if $|\beta^{(n+1)} - \beta^{(n)}| < \varepsilon \approx 10^{-8}$

end for

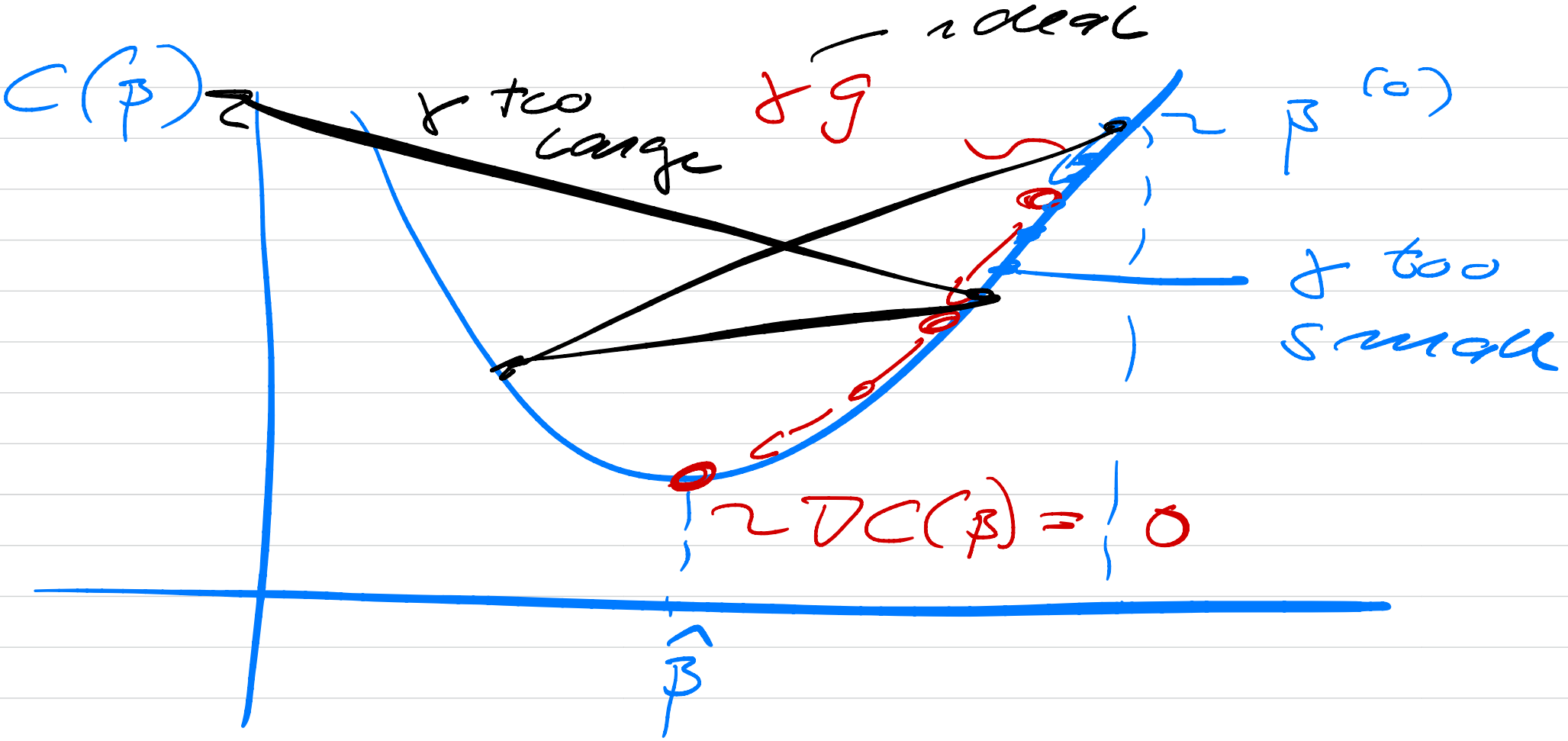
For ordinary least squares

$H = X^T X$ (no dependence on β)

Logistic regression

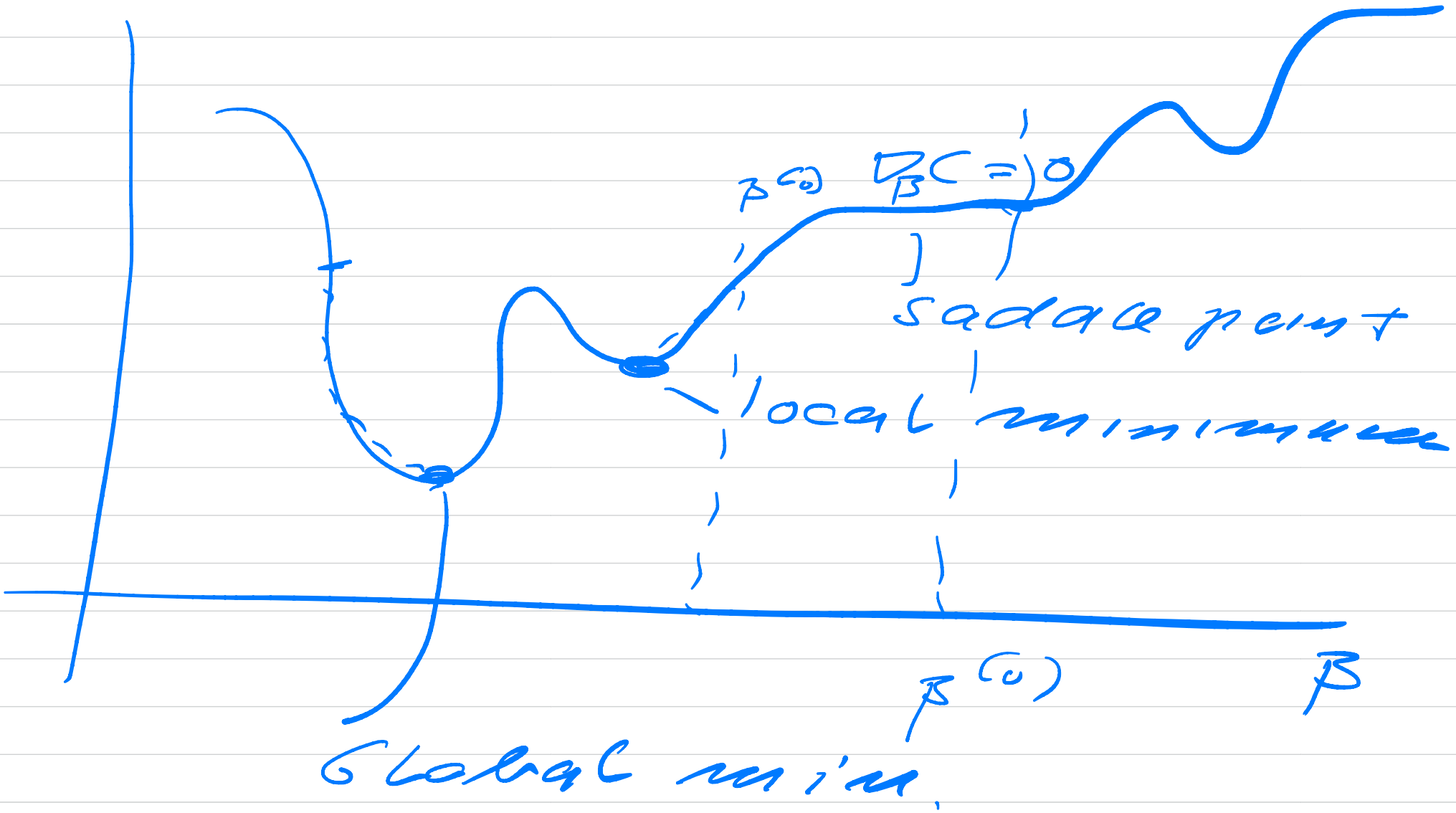
$H = X^T W X$

↑ implicit dependence on β



Goodfellow et al, chapters 4 and 8

General problem



steepest descent

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

(b: known)

$$\frac{\partial f(x)}{\partial x} = 0 \quad Ax - b \Rightarrow$$

$$Ax = b$$

solve iteratively

Define residual

$$r = b - Ax$$

start with guess r_0 and x_0

exact solution when $r = 0$

$x_0 = 0$, $r_0 = +b$

и далее

$$r_{k+1} = b - Ax_{k+1}$$

$$x_{k+1} = x_k + \alpha_k r_k$$

$$r_{k+1} = b - A(x_k + \alpha r_k)$$

$$= \underbrace{(b - Ax_k)}_{r_k} - \alpha_k A r_k$$

$$r_{k+1} = r_k - \alpha_k A r_k$$

$$r_{k+1} = 0 \quad \Rightarrow$$

$$r_k = \alpha_k A r_k \quad | \times r_k^T$$

$$\boxed{\alpha_k} = \frac{r_k^T r_k}{r_k^T A r_k} \quad | \; x_k - \alpha_k \underline{g_k}$$