9.3. ANDERE REGRESSIONSVERFAHREN

Lit. Murphy: Machine Learning – A probabilistic view

i-te Zeile der Design-Matrix A

Least squares Schätzer: $\underbrace{\sum_{i=1}^{n} |Y_i - w^T X_i|^2}_{} \overset{!}{=} \min$

Verlustfunktion = RSS (residual sum of squares)

PROBLEM 1: ROBUSTHEIT

Der LSE ist anfällig für Ausreißer.

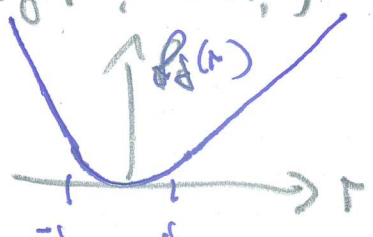
ALTERNATIVE 1 minimiere ℓ^1-Norm $\sum_{i=1}^{n} |Y_i - w^T X_i|$

$\hat{=}$ MLE im Laplace-Modell $f_\vartheta(y_i|x_i) \propto e^{-|y_i - w^T x_i|}$

Minimierung ist lineares Optimierungsproblem

ALTERNATIVE 2 minimiere Huber loss $\sum_{i=1}^{n} f_\delta(Y_i - w^T X_i)$

wobei $f_\delta(r) := \begin{cases} r^2/2 & \text{für } |r| \le \delta \\ \delta r - \frac{\delta^2}{2} & \text{für } |r| > \delta \end{cases}$

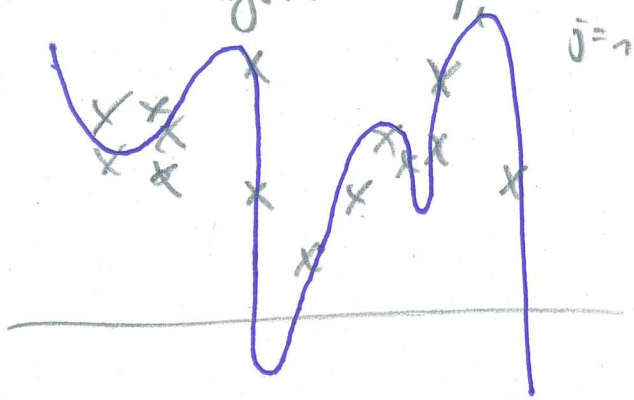


Vorteil: schnelle Optimierung durch numerische Verfahren (quasi-Newton)

PROBLEM 2: OVERFITTING

Bsp. Polynomielle Regression $y_i = \sum_{j=1}^{k} w_j x_i^j + \varepsilon_i$

k zu groß:



LSE hat zu viele große Gewichte w_j

ALTERNATIVE: Ridgeregression: $\lambda > 0$ fest

minimiere $$J(w) := \frac{1}{n} \underbrace{\sum_{i=1}^{n} |Y_i - w^T X_i|^2}_{RSS} + \underbrace{\lambda \|w\|^2}_{\text{penalty}}$$

bestraft große Gewichte

bzw. äquivalent: $B > 0$ fest

minimiere RSS unter Nebenbedingung $\|w\| \leq B$

λ groß $\longleftrightarrow$ B klein

$$\nabla J(w) = -\frac{2}{n} A^T Y + \frac{2}{n} A^T A w + 2\lambda w \overset{!}{=} 0$$

Min.

$$\rightsquigarrow \hat{w}_{ridge} = (\lambda I_d + A^T A)^{-1} A^T Y$$

λ größer $\Rightarrow$ mehr Regularität

Zudem numerisch vorteilhaft: Für $\lambda > 0$ einfache Berechnung von $\hat{\omega}_{ridge}$ via QR-Zerlegung

PROBLEM 3: HOCHDIMENSIONALER PARAMETERRAUM

z.B. $d > n \Rightarrow \mathrm{Rang}(A) < d \Rightarrow \ker(A) \neq \{0\}$

$\Rightarrow A\omega = \Pi Y$ hat keine eindeutige Lösung ω

ZIEL: SPARSITY Finde Schätzer $\hat{\omega}$ mit $\hat{\omega}_k = 0$ für "viele" k.

$\rightsquigarrow$ Identifikation der relevanten Einflussfaktoren

ANSATZ 1: Hard sparsity $\|\omega\|_0 := \#\{j \in \{1,\dots,d\} : \omega_j \neq 0\}$

$s \in \{1,\dots,d\}$ fest

Minimiere $\|Y - A\omega\|^2$ unter NB $\|\omega\|_0 \leq s$

bzw. Minimiere $J(\omega) := \underbrace{\|Y - A\omega\|^2}_{RSS} + \underbrace{\lambda \|\omega\|_0}_{\ell^0\text{ penalty}}$, $\lambda > 0$ fest

PROBLEME:

- $w \mapsto \|w\|_0$ unstetig, nicht konvex

  $\rightsquigarrow$ Optimierung von $\mathcal{J}$ ist aufwändig
- Größe der Gewichte wird nicht berücksichtigt

ANSATZ 2: Soft sparsity, ℓ^1 Regularisierung

$$\|w\|_0 \rightsquigarrow \|w\|_1 = \sum_{j=1}^{d} |w_j|$$

Lipschitz-stetig und konvex!

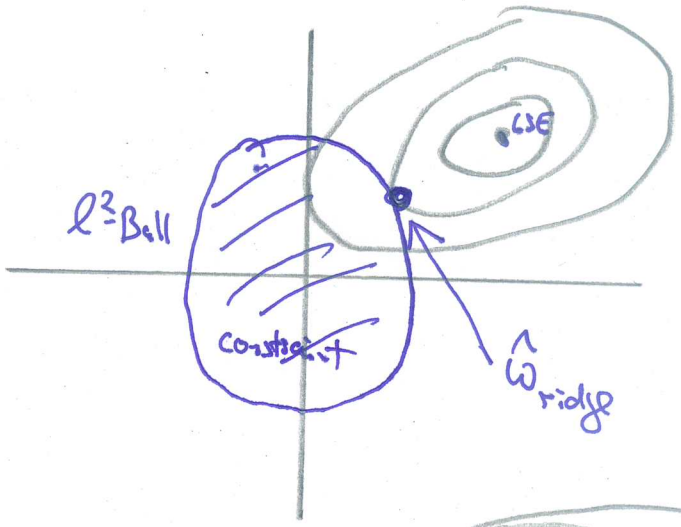
Minimiere $\|Y - Aw\|^2$ unter NB $\|w\|_1 \leq B$.

$\rightsquigarrow$ LASSO (least absolute shrinkage <u>and</u> selection operator)

konvexes Optimierungsproblem.

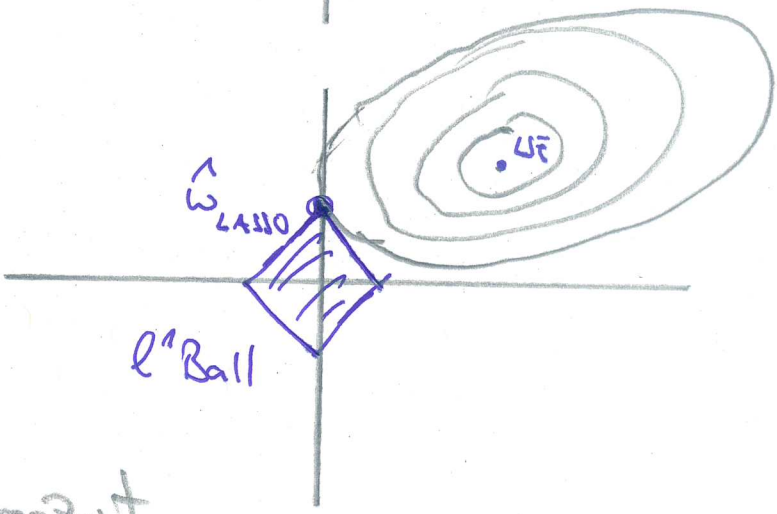
Ridge regression

$\|w\|_2 \leq B$



LASSO

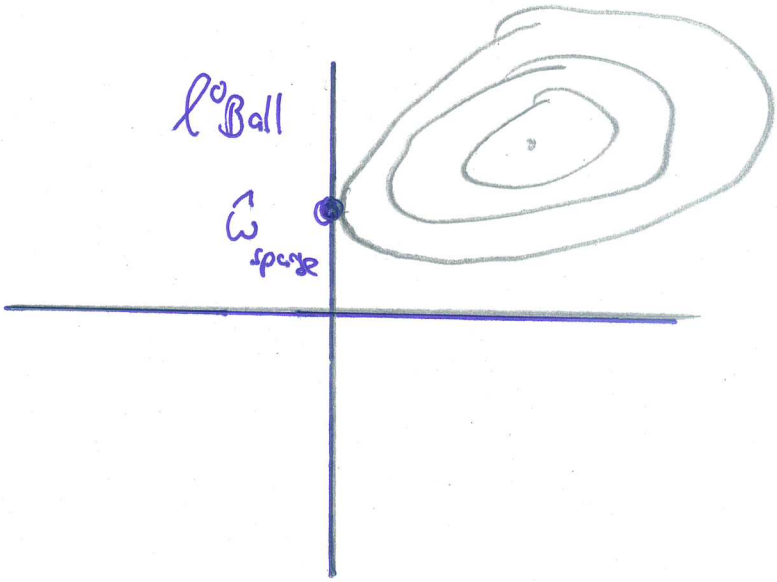
$\|w\|_1 \leq B$



Minimierer bevorzugt Ecken/Kanten/Hyperflächen des ℓ^1Balls, da diese mehr hervorstehen

Sparse regression

$\|w\|_0 \leq B$



PROBLEM 4: Binäre Response-Variable $Y_i \in \{0,1\}$

$\rightsquigarrow$ Classification

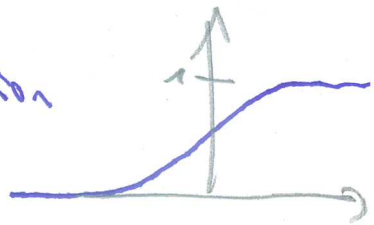
MODELL: Logistische Regression

$Y_i \sim \text{Bernoulli}(p_i)$ bedingt unabh. geg. X

$p_i = \text{sigm}(\omega^T X_i)$

$$\text{sigm}(x) = \frac{1}{1+e^{-x}}$$

logistische Funktion
Sigmoid



$$\underbrace{\log\left(\frac{\overbrace{p_i}^{\text{Chancen}}}{1-p_i}\right)}_{\text{logit-Funktion}} = \omega^T X_i$$

$$L(\omega; Y) = \prod_{i=1}^{n} \text{sigm}(\omega^T X_i)^{Y_i} \left(1-\text{sigm}(\omega^T X_i)\right)^{1-Y_i}$$

Berechnung des MLE z.B. via Newton-Verfahren

$\rightsquigarrow$ Reweighted Least Squares Algorithm