Sparse Recovery II

LASSO : $\min\limits_{w} \frac{1}{2} \| y - \Phi(x) \cdot w \|^2 + \alpha \cdot \| w \|_1$

Modellparameter $\in \mathbb{R}^{M+1}$

"least absolute shrinkage and selection operator"

Daten der abhängigen Größe $y = \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} \in \mathbb{R}^n$

approximierende Funktionen $\Phi(x) = \begin{pmatrix} \varphi^T(x^1) \\ \vdots \\ \varphi^T(x^n) \end{pmatrix} \in \mathbb{R}^{n \times (M+1)}$

Datenunabhängiger Größen $\in \mathbb{R}^p$

Bayes - Interpretation:

- $p(y^i; w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y^i - \varphi^T(x^i) \cdot w)^2}{2 \cdot \sigma^2}} \Leftarrow y^i \sim \mathcal{N}(\varphi^T(x^i) \cdot w, \sigma^2)$

$i = 1, \dots, n$

Fehlerdichte

normalverteilt:
- $E(y^i) = \varphi^T(x^i) \cdot w$
- $Var(y^i) = \sigma^2$.

"mehr Maß auf Erwartungswert und → "Null" oder "Groß" wahrscheinlich Tail"

- $p(w_j) = \frac{1}{2\tau} e^{-\frac{|w_j|}{\tau}} \Leftarrow w_j \sim Lap(0, \tau)$

$j = 0, \dots, M$

Priordichte

Laplace-verteilt:
- $E(w_j) = 0$
- $Var(w_j) = 2\tau^2$.

ü $\|\cdot\|_2^2$ - Regularisierung

$\underbrace{L(w)}_{\sim \text{ Bayes a-posteriori}} = \prod_{i=1}^{n} p(y^i; w) \cdot \prod_{j=0}^{M} p(w_j) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \cdot e^{-\frac{\|y - \Phi(x) \cdot w\|^2}{2\sigma^2}} \cdot \frac{1}{(2\tau)^{M+1}} \cdot e^{-\frac{\|w\|_1}{\tau}}$

iid-Messungen von Daten

$\max\limits_{w} \ln L(w) = \underbrace{\ln \left(\frac{1}{(\sqrt{2\pi\sigma^2})^n} \cdot \frac{1}{(2\tau)^{M+1}} \right)}_{\text{Konstante}} - \frac{\|y - \Phi(x) \cdot w\|^2}{2\sigma^2} - \frac{\|w\|_1}{\tau}$

Maximum Likelihood

$\longrightarrow \min\limits_{w} \underbrace{\frac{1}{2} \| y - \Phi(x) \cdot w \|^2}_{} + \underbrace{\frac{\sigma^2}{\tau} \|w\|_1}_{}$

$\frac{\sigma^2}{\tau} = \alpha$ - Verhältnis zwischen σ^2 und τ

"Gewichtung der Dünn-Besetztheit"

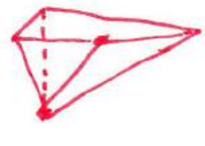
$F(w) := f(w) + g(w)$

$f(w)$:
- konvex
- differenzierbar

Approximation

$g(w)$:
- konvex
- nicht-differenzierbar

Einfache Funktion



Idee zur iterativen Berechnung von w^*

1. $\underbrace{\text{Modell}}_{\text{Obere Schranke}}$ für $f(w) = \frac{1}{2}\|y - \Phi \cdot w\|^2$

$$f(w) = f(v) + \underbrace{\int_0^1 f'(v + \tau(w-v))\, d\tau}_{\text{Hauptsatz der Differential- und Integralrechnung: } f(v+\tau(w-v))\big|_0^1 = f(w) - f(v)} =_{\text{Kettenregel}}$$

$$= f(v) + \int_0^1 \nabla^T f(v+\tau(w-v)) \cdot (w-v)\, d\tau$$

$$= f(v) + \nabla^T f(v) \cdot (w-v) + \int_0^1 \underbrace{\left[\nabla f(v+\tau(w-v)) - \nabla f(v)\right]^T (w-v)}_{\underset{\text{Cauchy-Schwarz } |a^T b| \le \|a\| \cdot \|b\|}{\le} \|\nabla f(v+\tau(w-v)) - \nabla f(v)\| \cdot \|w-v\|}\, d\tau$$

Lipschitz-Konstante > 0.

Annahme: $\|\nabla f(u) - \nabla f(v)\| \le L \cdot \|u - v\| \quad \forall u, v$

"Gradient von f ist Lipschitz-stetig"

$$\le L \cdot \|v + \tau(w-v) - v\| \cdot \|w - v\|$$

$$= L\, \tau\, \|w - v\|^2.$$

$$\le f(v) + \nabla^T f(v) \cdot (w-v) + \int_0^1 L \cdot \tau \cdot \|w-v\|^2\, d\tau =$$

$$= f(v) + \nabla^T f(v) \cdot (w-v) + L \cdot \|w-v\|^2 \cdot \underbrace{\int_0^1 \tau\, d\tau}_{= \frac{\tau^2}{2}\big|_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}}$$

$$= \underbrace{f(v) + \nabla^T f(v)(w-v) + \frac{L}{2}\|w-v\|^2}_{=: \tilde{f}(w,v) \text{ - Modell für } f}$$

$$\boxed{f(w) \le \tilde{f}(w,v) \quad \forall v}$$

Zur Überprüfung der Annahme:

$$\nabla f(w) = \nabla \left[\frac{1}{2}\left(y^T y - 2y^T \cdot \Phi \cdot w + w^T \cdot \Phi^T \cdot \Phi \cdot w\right)\right] = -\Phi^T \cdot y + \Phi^T \Phi \cdot w = \Phi^T \cdot (\Phi \cdot w - y)$$

$$\|\nabla f(u) - \nabla f(v)\| = \|\Phi^T(\Phi u - \cancel{y}) - \Phi^T(\Phi v - \cancel{y})\| = \|\Phi^T \Phi (u - v)\| =$$

$$= \|\Phi^T \cdot \Phi \underbrace{\frac{(u-v)}{\|u-v\|}}_{=: w \text{ mit } \|w\| = 1}\| \cdot \|u - v\| \le \underbrace{\max_{\|w\|=1} \|\Phi^T \cdot \Phi \cdot w\|}_{= \lambda_{max}(\Phi^T \Phi)} \cdot \|u - v\|$$

größter Eigenwert

$$\underbrace{\boxed{L := \lambda_{max}(\Phi^T \Phi) = \|\Phi\|^2.}}_{\text{Lipschitz-Konstante}}$$

Spektralnorm von Φ

$$\|\Phi\| := \max_{\|w\| \le 1} \|\Phi \cdot w\| = \sqrt{\lambda_{max}(\Phi^T \Phi)}$$

ü

2. Modell minimierung $\underbrace{f(w)+g(w)}_{=F(w)} \le \underbrace{\tilde{\tilde{F}}(w,v)+g(w)}_{\text{obere Schranke } = \tilde{F}(w,v)} \quad \forall v$

$$\min_w \; \cancel{f(v)} + \nabla^T f(v)\cdot(w-\cancel{v}) + \frac{L}{2}\|w-v\|^2 + \alpha\|w\|_1$$

"unabhängig von w"

$$\min_w \; \underbrace{\sum_{j=0}^{M} \nabla_j f(v)\cdot w_j + \frac{L}{2}(w_j-v_j)^2 + \alpha|w_j|}_{\text{minimiere bzgl. } w_j \text{, da separable Funktion}}$$

j - fest:

a) $w_j \ge 0 \Rightarrow \min_{w_j} \nabla_j f(v)\cdot w_j + \frac{L}{2}(w_j - v_j)^2 + \alpha\cdot w_j$

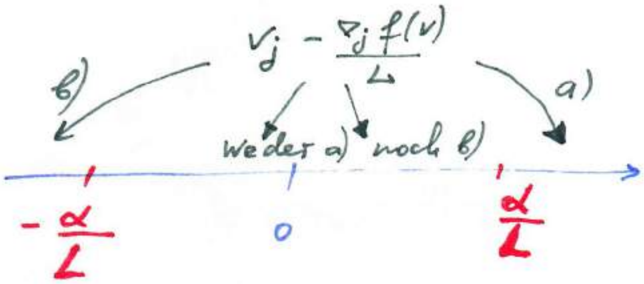
Optimalität: $\nabla_j f(v) + L(w_j - v_j) + \alpha = 0$

$$\Rightarrow w_j^a = \left[v_j - \frac{\nabla_j f(v)}{L}\right] - \frac{\alpha}{L}$$

b) $w_j < 0 \Rightarrow \min_{w_j} \nabla_j f(v)\cdot w_j + \frac{L}{2}(w_j-v_j)^2 - \alpha\cdot w_j$

Optimalität: $\nabla_j f(v) + L(w_j - v_j)^2 - \alpha = 0$

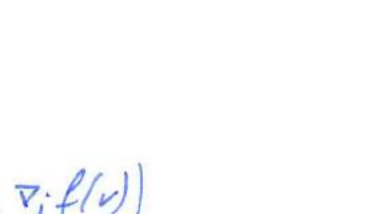
$$w_j^b = \left[v_j - \frac{\nabla_j f(v)}{L}\right] + \frac{\alpha}{L}$$



$$\Rightarrow w_j = \left[\left|v_j - \frac{\nabla_j f(v)}{L}\right| - \frac{\alpha}{L}\right]_+ \cdot \mathrm{sgn}\left(v_j - \frac{\nabla_j f(v)}{L}\right)$$

Lösung der Modellminimierung

Shrinkage / Soft-thresholding: $T_\beta(z) = \left[|z| - \beta\right]_+ \cdot \mathrm{sgn}(z)$



Schrittweite ↓ ← Gradientenschritt

$$\Rightarrow W = T_{\frac{\alpha}{L}}\left(V - \frac{1}{L}\cdot\underbrace{\nabla f(v)}_{=\Phi^T(\Phi\cdot v - y)}\right)$$

komponentenweise

↑ Thresholding

"kleine Einträge werden zu Null gesetzt"

⤳ Sparsity!

⟶ Iteration $\quad V := w(t)$ und $w := w(t+1)$

Iterative Shrinkage-Thresholding Algorithm (ISTA)

$$w(t+1) := T_{\frac{\alpha}{L}}\left(w(t) - \frac{1}{L}\cdot \Phi^T \cdot \left(\Phi \cdot w(t) - y\right)\right)$$

(t+1)-te Iterierte

t-te Iterierte

Konvergenz:

$$F(w(t+1)) \underset{\text{1. obere Schranke}}{\le} \tilde{F}(w(t+1), \overset{v}{\overset{\shortparallel}{w(t)}}) \underset{\text{2. Modellminimierung}}{=} \min_{w} \tilde{F}(w, w(t))$$

$$\underset{\min_w}{\le} \tilde{F}(w(t), w(t)) \underset{\text{einsetzen}}{=} F(w(t))$$

$\Rightarrow$ $F(w(t))$ ist monoton fallend $\Rightarrow$ Konvergenz gegen $\underbrace{\text{lokale}}_{=\text{globale}}$ Minima

= globale, da konvexes Optimierungsproblem

Ohne Beweis (analog):

$$F(w(t)) - F(w^*) \le \frac{\alpha \cdot L \cdot \overbrace{\| w(0) - w^* \|^2}^{\text{Abstand von der 0-ten Iterierten zur Lösung}}}{2 \cdot t} \approx \underbrace{O\left(\frac{1}{t}\right)}_{\text{Konvergenzrate ist sublinear}}$$

t-te Iterierte   Lösung des Optimierungsproblems