Zur Aufgabe 10.4

$U \subseteq \mathbb{R}^n$ k-dim. UVR, $x \in \mathbb{R}^n$ zufällig, $x_i \sim N(0, \frac{1}{n})$.

Was ist $E\left[\min_{u \in U} \|x-u\|^2\right]$, wenn $n \to \infty$?

Beh.: $E\left[\min_{u \in U} \|x-u\|^2\right] = \frac{n-k}{n} \xrightarrow[n\to\infty]{} 1$.

Bew.: Sei $u_1, \dots, u_k$ eine ONB von U und $u_{k+1}, \dots, u_n$ eine ONB von $U^\perp$. Dann ist

$$\min_{u \in U} \|x-u\|^2 = \sum_{i=k+1}^{n} (u_i^T x)^2 .$$

Also: $$E\left[\min_{u \in U} \|x-u\|^2\right] = \left(\frac{1}{\sqrt{\frac{2\pi}{n}}}\right)^n \int_{\mathbb{R}^n} \sum_{i=k+1}^{n} (u_i^T x)^2 \exp\left(-\frac{n}{2}\sum_{j=1}^{n} x_j^2\right) dx$$

[Variablentransformation: Ersetze x durch $A^{-1}x$,

mit $A = \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \in O(n)$.]

$$= \left(\frac{1}{\sqrt{\frac{2\pi}{n}}}\right)^n \int_{\mathbb{R}^n} \sum_{i=k+1}^{n} x_i^2 \exp\left(-\frac{n}{2}\sum_{j=1}^{n} x_j^2\right)$$

$$= E\left[\sum_{i=k+1}^{n} x_i^2\right] = \frac{n-k}{n}.$$

Klassifikation:

$$X = \{x_1, \dots, x_M\}, \quad Y = \{y_1, \dots, y_N\} \subseteq \mathbb{R}^n.$$

X und Y können durch eine affin-lineare Fkt.

$f(x) = a^T x - \beta$ getrennt werden $\iff$

$$0 < t^* = \max t$$
$$a^T x_i - \beta \geq t \quad , i = 1, \dots, M$$
$$a^T y_i - \beta \leq -t \quad , i = 1, \dots, N$$
$$\|a\| \leq 1$$
$$= \min \frac{1}{2} \|x - y\|$$
$$x \in \operatorname{conv} X,$$
$$y \in \operatorname{conv} Y.$$

In realistischen Szenarien können X und Y nicht mit einer affin-linearen Fkt. getrennt werden.

Dann werden sog. "support vector machines" (SVM) verwendet.

Ersetze in der Originalformulierung

$$\begin{aligned} a^T x_i - \beta &\geq 1 \\ a^T y_i - \beta &\leq -1 \end{aligned} \quad \text{durch} \quad \begin{aligned} a^T x_i - \beta &\geq 1 - u_i \\ a^T y_i - \beta &\leq -(1 - v_i) \end{aligned} \qquad (*)$$

für $u_i \geq 0$, $v_i \geq 0$.

Falls $\mu = \nu = 0$, dann haben wir die Originalformulierung.

Falls μ, ν genügend groß ist, können die Bed. (*) immer erfüllt werden.

Heuristik: Minimiere die ℓ_1 - Norm des Vektors (μ, ν)

Ergibt lineares Programm:

$$\min \sum_{i=1}^{M} \mu_i + \sum_{i=1}^{N} \nu_i$$

$$\mu_i \geq 0, \quad i = 1, \dots, M,$$

$$\nu_i \geq 0, \quad i = 1, \dots, N,$$

$$a^T x - \beta \geq 1 - \mu_i, \quad i = 1, \dots, M,$$

$$a^T x - \beta \leq -(1 - \nu_i), \quad i = 1, \dots, N.$$

Alternative: Minimiere ℓ_1 - Norm von (μ, ν) und gleichzeitig maximiere die Breite des Streifens $\{x : -1 \leq a^T x - \beta \leq 1\}$, der durch $\frac{2}{\|a\|}$ gegeben ist.

Für $\gamma > 0$ löse das konvexe, quadratische Programm

$$\min \|a\| + \gamma \left(\sum_{i=1}^{M} \mu_i + \sum_{i=1}^{N} \nu_i \right)$$

$$a^T x_i - \beta \geq 1 - \mu_i$$

$$a^T y_i - \beta \leq -(1 - \nu_i)$$

$$\mu_i \geq 0, \quad \nu_i \geq 0.$$

"ℓ_1-Norm soft-margin SVM"

§ 2 Nicht-lineare Klassifikation

<u>Idee</u>: Verwende nicht-lineare Einbettung von $X, Y \subseteq \mathbb{R}^n$ in m-dim. VR, $m >> n$.

$$\varphi : \mathbb{R}^n \to \mathbb{R}^m, \quad \text{z. B.} \quad \underset{\mathbb{R}^2}{\underset{\cap}{\varphi(x_1, x_2)}} = \underset{\mathbb{R}^5}{\underset{\cap}{(x_1^2, x_1 x_2, x_2^2, x_1, x_2)}}.$$

Dann können nicht-lineare trennende Hyperflächen im $\mathbb{R}^n$ durch lineare Hyperebenen im $\mathbb{R}^m$ approximiert werden.

<u>Problem</u>: Falls m sehr groß ist, ist das ℓ_1-norm soft-margin SVM aufwendig zu lösen.

<u>Lösung</u>: Dualisiere und verwende den "<u>Kernel trick</u>".

Das Duale ist:

$$\max \sum_{i=1}^{M} \lambda_i + \sum_{i=1}^{N} \mu_i$$

$$\left\| \sum_{i=1}^{M} \lambda_i x_i - \sum_{i=1}^{N} \mu_i y_i \right\| \leq 1$$

$$\lambda_i \in [0, S], \quad i = 1, \dots, M$$

$$\mu_i \in [0, S], \quad i = 1, \dots, N,$$

$$\sum_{i=1}^{M} \lambda_i = \sum_{i=1}^{N} \mu_i$$

Wenn man die Einbettung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ verwendet, muß man die Bedingung

$$\left\| \sum_{i=1}^{n} \lambda_i \varphi(x_i) - \sum_{i=1}^{N} \mu_i \varphi(y_i) \right\|^2 \leq 1$$

in den Griff bekommen.

Der "<u>Kernel-Trick</u>": Diese Bedingung hängt nur von der Gram-Matrix $K : (X \cup Y) \times (X \cup Y) \to \mathbb{R}$ ab:

$$K(x_i, y_j) = \varphi(x_i)^T \varphi(y_j) .$$

Falls man K effizient berechnen kann, dann ist die Dim. m der Einbettung irrelevant.

Die Fkt., die die Gram-Matrix beschreibt heißt <u>Kern</u>.

Beliebte Kerne sind (Wahl ist extrem anwendungsabh.)

(a) $K(x_i, x_j) = x_i^T x_j$ <u>linear</u>

(b) $K(x_i, x_j) = (x_i^T x_j)^d$ <u>d-te Potenz</u>

(c) $K(x_i, x_j) = \exp\left(-\frac{1}{2} \|x_i - x_j\|^2 / \sigma^2\right)$

<u>radiale Basisfkt</u> (oft verwendet)