Blatt 4: Regression + SVM

Datenerzeugung: Erzeugen Sie 1D-Daten einer Funktion $f(x) : x \in \mathbb{R} \to \mathbb{R}$ und visualisieren Sie diese. Die Daten sollten auch Lücken und Ausreißer enthalten.

Aufgabe 1: Lineare-Regression

Schätzen Sie die den Daten zugrundeliegende Funktion mittels einer linearen Regression (mit den Formeln aus dem Skript) und visualisieren Sie diese zusammen mit den Daten.
Kommentieren Sie die Ergebnisse.

Aufgabe 2: Lineare-Regression mit Polynomen

Schätzen Sie die den Daten zugrundeliegende Funktion mittels einer linearen Regression mit einem vollständigen Polynom vom Grad $\nu = 2, 3, 5$ usw. und visualisieren Sie diese.
Es soll nach dem Skript vorgegangen werden, d.h. keine Bibliotheken für Matlab/Python verwendet werden.
Kommentieren Sie die Ergebnisse.

Aufgabe 3: Support-Vektor-Maschinen

Large–Margin Klassifikation: Die wesentlichen Aussagen

- Die Supportvektor–Maschine ist von Natur aus ein Klassifikator für $c = 2$ Klassen.
- Das Optimierungskriterium bestimmt eine Entscheidungsgrenze, welche den Rand zwischen den Klassen maximiert.
- Im nicht–separablen Fall werden neben dem Rand mit einem Gewichtsparameter C auch die Summe positiver Schlupfvariable ξ_i optimiert, die eine individuelle "Aufweichung" der Nebenbedingungen erlauben. Fehlklassifizierte Muster der Stichprobe sind somit möglich, werden allerdings bestraft.
- Nichtlineare Trennfunktionen können bereits mit dem gleichen Optimierungsansatz mittels der nichtlinearen Transformation des Merkmalraumes mit Kernfunktionen berechnet werden. Dadurch wird das Skalarprodukt im Merkmalsraum durch das Skalarprodukt im transformierten Merkmalsraum ersetzt.
- Die Lösung wird sparsam dargestellt mittels einer Auswahl aus den Trainingsvektoren, den sogenannten **Support-Vektoren**.
- Unter "konvexer quadratischer Programmierung" versteht man die Minimierung einer konvexen quadratischen Kostenfunktion (d.h. mit positiv-semidefiniter Matrix im quadratischen Term) unter linearen Nebenbedingungen (Gleichungen und/oder Ungleichungen), d.h. über einer konvexen Menge. Ein solches Optimierungsproblem ist in der Trainigsphase zur Berechnung eines Supportvektor–Klassifkators zu lösen.
- Folgende Begriffe der konvexen Optimierung sind dabei wichtig:
  - Konvexe Funktionen
  - Lagrangefunktion, Lagrangesche Multiplikatoren
  - KKT-Bedingungen, insbesondere auch Komplementaritätsbedingung
  - duale Funktion, duales Optimierungsproblem

  Bei der SVM–Klassifikation wird das duale Optimierungsproblem numerisch gelöst (Variablenvektor: α). Die primären Variablen $\mathbf{w}, b$ ergeben sich dann aus den KKT–Bedingungen.

  Literatur:
  Falls Sie ein Update zur nichtlinearen Optimierung unter Nebenbedingungen brauchen: Infos finden Sie entweder in Ihrem Mathematik-Skript, im Bronstein (unter Lagrange-Multiplikatorenmethode) oder in einem beliebigen Buch zu numerischer Analysis und/oder nichtlinearer Optimierung (z.B. Bertsekas, Nonlinear Programming). Für das Verständnis der Vorlesung reichen die paar Zeilen aus dem Bronstein bereits aus.

SVM-Klassifikation:

Diese Teilaufgabe ist **handschriftlich** zu bearbeiten.

Gegeben ist eine Stichprobe $\mathcal{X}$ einer vektoriellen Zufallsvariable $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2,\ i = 1, \ldots, 7$:

$$\mathcal{X} = \Big\{ \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \mathbf{x}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix},\ \mathbf{x}_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},\ \mathbf{x}_4 = \begin{pmatrix} 3 \\ 1 \end{pmatrix},\ \mathbf{x}_5 = \begin{pmatrix} 2 \\ 2.5 \end{pmatrix},\ \mathbf{x}_6 = \begin{pmatrix} 3 \\ 2.5 \end{pmatrix},\ \mathbf{x}_7 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \Big\}$$

Dabei gehören die $\mathbf{x}_i,\ i \in \{2, 3, 5\}$ zur Klasse 1, und die übrigen zur Klasse 2 (wie in Blatt-4).

Der SVM-Klassifikator lautet allgemein (die Übersichtsfolie aus der Vorlesung darf dabei als gegeben angenommen werden):

$$f(\mathbf{x}) = \operatorname{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b), \qquad \text{mit} \qquad \langle \mathbf{w}, \mathbf{x} \rangle = \sum_{\mathbf{x}_i \in SVs} \alpha_i y_i \langle \mathbf{x}, \mathbf{x}_i \rangle$$

In allen Teilaufgaben soll der „lineare“ Kernel $K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{y} + 1$ verwendet werden.

1. Wie lautet der Vektor $\mathbf{y}$ bestehend aus den $y_i,\ i = 1, \ldots, 7$ der Klassifikationskennungen für einen SVM-Klassifikator? (Muster der Klasse 1 werden mit $y = 1$ gekennzeichnet)
2. Zeichnen Sie das Problem.
   Handelt es sich um eine linear trennbare Problemstellung?
   Geben Sie zeichnerisch eine lineare Trennfunktion an, welche etwa den Rand maximiert.
3. Geben Sie zu Ihrer zeichnerischen Lösung passende Parameter $\mathbf{w}, b$ der linearen Trennfunktion an ($f(x) = \mathbf{w}^\top \mathbf{x} + b$).
4. Ein Paket zur numerischen Lösung des dualen Problems unter Verwendung des obigen Kernels ergibt:
   $$\boldsymbol{\alpha}^* = \begin{pmatrix} 0 & 8.23 & 0 & 0 & 0.95 & 8.23 & 0.95 \end{pmatrix}^\top$$
   Welche Support-Vektoren resultieren daraus?
5. Geben Sie unter Verwendung des obigen Kernels das Innenprodukt $\langle \mathbf{w}, \mathbf{x} \rangle$ an.
   Wie lautet der daraus resultierende Vektor für $\mathbf{w}^*$?
   Was fällt Ihnen auf, wenn Sie $\mathbf{w}^*$ mit Ihrem $\mathbf{w}$ aus Teilaufgabe (3) vergleichen?
6. Bestimmen Sie b aus der Komplementaritätsbedingung
   $$\alpha_i[(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) y_i - 1] = 0,\ \forall i$$
   Mitteln Sie dabei das Ergebnis über alle Support-Vektoren.
7. Klassifizieren Sie nun das Muster $\mathbf{x}_s = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ mittels der Funktion $f(\mathbf{x})$. Begründen Sie Ihre Entscheidung.

SVM-Regression

Erstellen Sie Daten einer 1D-Funktion, für die Sie eine SVM-Regression ansetzen und berechnen.

Bilden Sie sich damit eine Meinung zur Wirkung der Parameter $C,\ \sigma,\ \varepsilon$ und begründen Sie diese.

Es kann hier eine fremde Bibliothek zur numerischen Lösung eingesetzt werden.