Grundlagen des maschinellen Lernens (2023/24)

Aufgabenblatt 3

Abgabetermin: Freitag, 19. Januar 2024

Die Abgabe erfolgt durch Hochladen der Lösungen in ILIAS.

Die Lösung zu Aufgabe 1 soll als (möglichst mit LaTeX erzeugtes) PDF-File eingereicht werden, die Lösung zu Aufgabe 2 ist als Jupyter-Notebook einzureichen.

Aufgabe 1

Gegeben seien die folgendermaßen bestimmten vier Datenpunkte $\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),(\boldsymbol{x}_3,y_3),(\boldsymbol{x}_4,y_4)\}$ in $\mathbb{R}^2\times\{0,1\}$:

$$\begin{aligned}
\boldsymbol{x}_1 &= (\ \ 1,\ \ 2),\ \ y_1=1,\\
\boldsymbol{x}_2 &= (-2,\ \ 1),\ \ y_2=0,\\
\boldsymbol{x}_3 &= (-1,-2),\ \ y_3=1,\\
\boldsymbol{x}_4 &= (\ \ 2,-1),\ \ y_4=0.
\end{aligned}$$

Ziel der Aufgabe ist es, zu dem Datensatz einen SVM-Klassifikator mit polynomialem Kernel zu bestimmen, wobei die Kernelfunktion $\mathscr{K}$ durch die Merkmalstransformation $\phi:\mathbb{R}^2\to\mathbb{R}^4$, $\boldsymbol{x}=(x_1,x_2)\mapsto(x_1^2,x_2^2,x_1x_2,x_1x_2)$ gegeben sei.

(i) Drücken Sie $\mathscr{K}(\boldsymbol{x},\boldsymbol{x}')$ in Abhängigkeit von $\boldsymbol{x}^\mathsf{T}\boldsymbol{x}'$ aus (mit Angabe der Umformungsschritte) und bestimmen Sie die durch die Datenpunkte gegebene Matrix $\boldsymbol{K}:=(\mathscr{K}(\boldsymbol{x}_i,\boldsymbol{x}_j))_{i,j=1,2,3,4}$.

(ii) Ermitteln Sie die dualen Parameter $\alpha^*=(\alpha_1^*,\alpha_2^*,\alpha_3^*,\alpha_4^*)$ des gesuchten Modells, indem Sie das zugehörige duale Optimierungsproblem mit Nebenbedingungen (für den 'hard margin'-Fall) lösen. Wählen Sie Ihre Lösung so, dass sie einer möglichst kleinen Anzahl von Stützvektoren entspricht.

Hinweis: Leiten Sie den zu maximierenden Ausdrucks partiell nach den α_i's ab und setzen Sie die Ableitungen gleich 0. Beachten Sie zusätzlich die Nebenbedingungen.

Tipp: Sie können die Korrektheit Ihres Ergebnisses überprüfen, indem Sie die Parameter des gesuchten SVM-Modells mit Hilfe von Scikit-Learn unter Verwendung der Klasse `SVC` mit `kernel="poly"` ermitteln. (Beachten Sie dabei die korrekte Wahl aller Parameter des Modells. Außerdem werden möglicherweise andere Stützvektoren ausgewählt.)

(iii) Geben Sie den durch die dualen Parameter bestimmten linearen Prädiktor f des SVM-Modells in Abhängigkeit von $\boldsymbol{x}=(x_1,x_2)$ an. Berechnen Sie $f(0,1)$ und $f(1,0)$.

(iv) Berechnen Sie die primären Parameter $\boldsymbol{w}=(w_0,w_1,w_2,w_3,w_4)$ des SVM-Modells aus den ermittelten dualen Parametern. Zeigen Sie, dass jede mögliche Lösung für die dualen Parameter zum gleichen $\boldsymbol{w}$ führt.

Aufgabe 2

Betrachten Sie nochmals die in Aufgabe 1 von Aufgabenblatt 2 definierte Datenmenge. Teilen Sie die Datenpunkte wie dort beschrieben in eine Trainings- und eine Validierungsmenge auf.

Die Trainingsdaten sollen nun mittels SVM-Regression unter Einsatz des folgendermaßen definierten RBF-Kernels $\mathscr{K}$ modelliert werden:

$$\mathscr{K}(\boldsymbol{x},\boldsymbol{x}') \;=\; \exp\Big(-\left(\frac{\boldsymbol{x}-\boldsymbol{x}'}{b}\right)^2\Big)$$

Verwenden Sie dazu die Klasse `SVR` aus Scikit-Learn mit den Parametern `kernel="rbf"`, `C=100` und `epsilon=0.1`. (Beachten Sie, dass der zusätzliche Parameter `gamma` in geeigneter Weise aus dem Wert von b errechnet werden muss.)

(i) Ermitteln Sie denjenigen Wert für b, für den das berechnete Modell den geringsten Fehler auf den Validierungsdaten aufweist. Testen Sie dazu die ersten zweihundert Vielfachen von 0.01 als mögliche Werte für b. Verwenden Sie wieder die Funktion `mean_squared_error` aus Scikit-Learn zur Berechnung des Fehlers. Erzeugen Sie einen Plot des Fehlerwerts in Abhängigkeit von b.

(ii) Ermitteln Sie außerdem denjenigen Wert für b, für den der Fehler minimal ist, bei gleichzeitiger Minimierung der Anzahl der Stützvektoren. (Nutzen Sie dazu geeignete Attribute der Klasse `SVR`.) Erzeugen Sie einen Plot der Anzahl der Stützvektoren in Abhängigkeit von b.