MUSTERERKENNUNG, 10. SEMINAR – HINGE-LOSS

Aufgabe 1.

a) Sei $x \in \mathbb{R}$ und $y \in \mathbb{R}$ zwei reellwertige Variablen. Betrachten Sie die folgende Optimierungsaufgabe:

$$\begin{aligned} &|x| + |y| \to \min_{xy} \\ &\text{s.t. } \; ax + by \geq 1, \end{aligned}$$

mit reellwertigen Koeffizienten a und b. Geben Sie eine geometrische Interpretation der Aufgabe an und zeigen Sie, dass im Optimum eine der Variablen (d.h. entweder x oder y) gleich Null ist (bei beliebigen a und b).

b) Verallgemeinern Sie diese Aussage auf höhere Dimensionen, d.h. betrachten Sie die Aufgabe

$$\begin{aligned} &|x|_1 = \sum_i |x_i| \to \min_x \\ &\text{s.t. } \; \langle x, a \rangle \geq 1, \end{aligned}$$

mit $x, a \in \mathbb{R}^n$.

Aufgabe 2. Zur Klassifikation der Muster $x \in \mathbb{R}$ soll die Schwellwertentscheidung verwendet werden, d.h. $y = \operatorname{sign}(x - \theta)$ mit ± 1-Kodierung der Klassen und einem Schwellwert θ. Zum Anlernen des unbekannten Schwellwerts steht die folgende klassifizierte Lernstichprobe zur Verfügung: $L = \big((x^l, y^l) \ldots\big) = \big((0,-1),(3,+1),(6,-1)\big)$ – sie besteht aus drei Punkten und ist offensichtlich nicht separierbar.

a) Zeichnen Sie den Graph des Empirischen Risikos in Abhängigkeit von θ.

b) Um den optimalen Schwellwert zu finden, wird das Empirische Risiko durch Hinge-Loss ersetzt. Zeichnen Sie den Graph des Hinge-Losses in Abhängigkeit von θ. Vergleichen Sie die beiden Funktionen.

c) Konstruieren Sie eine Lernstichprobe derart, dass das Minimum des Hinge-Losses an solchen Stellen erreicht wird, an welchen das Empirische Risiko maximal ist.

Aufgabe 3. Bei der Vorlesung haben wir gezeigt, dass die Minimierung des Hinge-Losses (ergänzt mit einem quadratischen Regularizator) unter Umständen dem Maximum Margin Lernen entspricht (Folien 9,10). Jetzt diskutieren wir eine alternative Vorgehensweise.

Man betrachte das Maximum Margin Lernen, in dem die Fehlklassifikationen erlaubt (obwohl immer noch „nicht erwünscht“) sind. Man führt für jedes Lernbeispiel eine so genannte slack-Variable ξ_l ein, die die entsprechende Nebenbedingung in einem gewissen Sinn abschwächt. Gleichzeitig repräsentieren diese Variablen Fehler und müssen somit bestraft werden. Zusammenfassend lautet die Aufgabe (für lineare Klassifikatoren als Beispiel)

$$\begin{aligned} & ||w||^2 + \frac{C}{L}\sum_l \xi_l \to \min_{w,\xi} \\ & \text{s.t.} \quad y_l[\langle x_l, w\rangle - b] \geq 1 - \xi_l, \;\; \xi_l \geq 0 \;\; \forall l. \end{aligned}$$

Zeigen Sie, dass eine solche Formulierung dem Minimieren des Hinge-Losses äquivalent ist, wenn der letztere mit einem quadratischen Regularizator versehen ist.

Hinweis: Nutzen Sie den folgenden Trick: $\max(a,b) = \min \xi$, s.t. $\xi \geq a,\ \xi \geq b$