STRUCTURED MODEL LEARNING (WS2021/22)
SEMINAR 3

Assignment 1. Let $\mathcal{G} \subseteq [0,1]^{\mathcal{Z}}$ be a set of functions $g\colon \mathcal{Z} \to [0,1]$. Let $\mathcal{U}^m = \{z^1, \ldots, z^m\} \in \mathcal{Z}^m$ be drawn i.i.d. from $p(z)$. The Rademacher complexity of $\mathcal{G}$ w.r.t. the distribution $p(z)$ is

$$\hat{\mathcal{R}}_m(\mathcal{G}) = \mathbb{E}_{\mathcal{U}^m \sim p^m(z)} \mathbb{E}_{\sigma \sim \mathrm{Unif}\{-1,+1\}} \left[\sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i \, g(z_i) \right]$$

a) What is the minimal value of the Rademacher complexity?

b) What is the value of the Rademacher complexity when $\mathcal{G}$ contains just a single function, i.e. $|\mathcal{G}| = 1$?

c) What is the maximal value of the Rademacher complexity? What is the minimal number of functions in $\mathcal{G}$ to achieve the maximal value?

Assignment 2. Let $\{(\boldsymbol{x}^i, y^i) \in \mathbb{R}^n \times \{-1,+1\} \mid i = 1, \ldots, m\}$ be m points in $\mathbb{R}^n$ that are assigned into two classes. Assume that there exists an ellipse separating the points in the positive class from the points in negative class, i.e., there exists a positive definite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and scaler $r > 0$ such that

$$\begin{array}{lcll} \langle \boldsymbol{x}^i, \mathbf{A}\boldsymbol{x}^i \rangle & \geq & r^2\,, & \forall i \in \{j \in \{1, \ldots, m\} \mid y^i = +1\}\,, \\ \langle \boldsymbol{x}^i, \mathbf{A}\boldsymbol{x}^i \rangle & < & r^2\,, & \forall i \in \{j \in \{1, \ldots, m\} \mid y^i = -1\}\,. \end{array} \tag{1}$$

Show how to use the Perceptron algorithm to find $\mathbf{A}$ and r which satisfy the inequalities (1).

Assignment 3. Let $\mathcal{X} = \mathcal{A}^n$ be a set of input sequences and $\mathcal{Y} = \mathcal{B}^n$ a set of hidden sequences of length n which are defined over finite alphabets $\mathcal{A}$ and $\mathcal{B}$, respectively. Let $h\colon \mathcal{X} \to \mathcal{Y}$ be a prediction rule that for each $x \in \mathcal{X}$ returns a sequence $h(x) = (h_1(x), \ldots, h_n(x))$ obtained solving

$$h(x) = \operatorname*{arg\,max}_{(y_1, \ldots, y_n) \in \mathcal{B}^n} \left(\sum_{i=1}^n q(x_i, y_i) + \sum_{i=2}^n g(y_{i-1}, y_i) \right) \tag{2}$$

where $q\colon \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ and $g\colon \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ are quality fucntions describing compatibility between inputs and hidden states.

a) Show that (2) is a linear classifier.

b) Describe a dynamic programming algorithm which computes the output of the classifier (2) in time polynomial in the size of the input instances . How does the algorithm scale with $|\mathcal{A}|$, $|\mathcal{B}|$ and n ?

c) Describe an instance of Perceptron algorithm which learns the quality functions q and g from linearly separable examples $\{(x_1^j,\ldots,x_n^j,y_1^j,\ldots,y_n^j)\in\mathcal{A}^n\times\mathcal{B}^n \mid j=1,\ldots,m\}$.

Assignment 4. Consider a linear ordinal classifier $h\colon \mathbb{R}^n\to\{1,\ldots,Y\}$ defined by

$$h(\boldsymbol{x}) = 1 + \sum_{y=1}^{Y-1} [\![\langle \boldsymbol{w},\boldsymbol{x}\rangle \geq b_y]\!] \qquad (3)$$

and parameterized by a vector $\boldsymbol{w}\in\mathbb{R}^n$ and an increasing sequence of thresholds $b_1 < b_2 < \cdots < b_{Y-1}$. Let $\mathcal{T}^m = \{(\boldsymbol{x}^j,y^j)\in(\mathbb{R}^n\times\mathcal{Y}) \mid j=1,\ldots,m\}$ be a training set of examples. Describe a variant of the Perceptron algorithm which finds the parameters $\boldsymbol{w}\in\mathbb{R}^n$ and $b_y\in\mathbb{R}$, $y\in\{1,\ldots,Y-1\}$, such that the classifier (3) predicts all examples from $\mathcal{T}^m$ correctly provided such parameters exist.