STRUCTURED MODEL LEARNING (SS2015)
5.SEMINAR

Assignment 1. Let $h\colon \mathcal{X} \to \mathcal{Y}$ be a fixed classification strategy. Let $\mathcal{T} = \{(\boldsymbol{x}^i, \boldsymbol{y}^i) \in \mathcal{X} \times \mathcal{Y} \mid i \in \{1,\ldots,m\}\}$ be a training set drawn from i.i.d. random variables with distribution $p(\boldsymbol{x},\boldsymbol{y})$. A probability that the expected risk $R(h) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p}[\ell(\boldsymbol{y}, h(\boldsymbol{x})]$ deviates from the empirical risk $R_{\mathcal{T}}(h) = \frac{1}{m}\sum_{i=1}^m \ell(\boldsymbol{y}^i, h(\boldsymbol{x}^i))$ by at least $\varepsilon > 0$ can be bounded by the Hoeffding's inequality

$$\mathcal{P}_{\mathcal{T}\sim p^m}\left(\Big|R(h) - R_{\mathcal{T}}(h)\Big| \geq \varepsilon\right) \leq 2\exp\left(\frac{-2m\,\varepsilon^2}{l_{\max}}\right)$$

where l_{max} is the maximal value of the loss function $\ell\colon \mathcal{Y}\times\mathcal{Y} \to [0, l_{max}]$. Prove that for a finite hypothesis space $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{Y}}$, i.e. $|\mathcal{H}| < \infty$, the following inequality holds

$$\mathbb{P}_{\mathcal{T}\sim p^m}\left(\max_{h\in\mathcal{H}}\Big|R(h) - R_{\mathcal{T}}(h)\Big| \geq \varepsilon\right) \leq 2|\mathcal{H}|\exp\left(\frac{-2m\,\varepsilon^2}{l_{max}}\right)$$

Hint: A similar bound was proven in Lecture 6, section 5 for the likelihood function. Use the same procedure.

Assignment 2. Consider multi-class linear classifier $h\colon \mathbb{R}^n \to \mathcal{Y} = \{1,\ldots,K\}$ defined by

$$h(\boldsymbol{x}; \boldsymbol{w}_1,\ldots,\boldsymbol{w}_K) = \operatorname*{argmax}_{y\in\mathcal{Y}} \langle \boldsymbol{x}, \boldsymbol{w}_y\rangle$$

where $\boldsymbol{w} = (\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k) \in \mathbb{R}^{n\times K}$ are parameters. Derive a variant of the Perceptron algorithm to learn the parameters $\boldsymbol{w}$ from linear separable examples $\{(\boldsymbol{x}^i, \boldsymbol{y}^i) \in \mathbb{R}^n \times \mathcal{Y} \mid i \in \{1,\ldots,m\}\}$.

Assignment 3. Consider joint distribution

$$p_{\boldsymbol{q},\boldsymbol{g}}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{q},\boldsymbol{g})}\exp\left(\sum_{v\in\mathcal{V}} q_v(x_v, y_v) + \sum_{\{v,v'\}\in\mathcal{E}} g_{vv'}(y_v, y_{v'})\right) = \frac{1}{Z(\boldsymbol{q},\boldsymbol{g})}\exp f(\boldsymbol{x},\boldsymbol{y};\boldsymbol{q},\boldsymbol{g})$$

Show that the optimal (Bayes) classifier minimizing the expected risk with 0/1-loss, i.e.

$$R(h) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p_{\boldsymbol{q},\boldsymbol{g}}} [\![\boldsymbol{y} \neq h(\boldsymbol{x})]\!]$$

is the max-sum classifier

$$h(\boldsymbol{x};\boldsymbol{q},\boldsymbol{g}) \in \operatorname*{argmax}_{\boldsymbol{y}\in\mathcal{Y}^{\mathcal{V}}} f(\boldsymbol{x},\boldsymbol{y};\boldsymbol{q},\boldsymbol{g})$$

For the same distribution $p_{q,g}(\boldsymbol{x},\boldsymbol{y})$ derive the optimal classifier $h\colon \mathcal{X}^{\mathcal{V}} \to \mathcal{Y}^{\mathcal{V}}$ minimizing the expected risk with the Hamming distance used as the loss, i.e

$$R(h) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p_{\boldsymbol{q},\boldsymbol{g}}}\left[\sum_{v\in\mathcal{V}} [\![y_v \neq h_v(\boldsymbol{x})]\!]\right]$$

Assignment 4. The LP relaxation of the max-sum problem reads

$$\boldsymbol{\mu}^* = \operatorname*{argmax}_{\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{V}||\mathcal{Y}|+|\mathcal{E}||\mathcal{Y}|^2}} \left[\sum_{v\in\mathcal{V}} \sum_{y\in\mathcal{Y}} \mu_v(y) q_v(x_v, y_v) + \sum_{\{v,v'\}\in\mathcal{E}} \sum_{(y,y')\in\mathcal{Y}^2} \mu_{v,v'}(y,y') g_{v,v'}(y,y') \right]$$

subject to

$$\sum_{y'\in\mathcal{Y}} \mu_{v,v'}(y,y') = \mu_v(y), \{v,v'\}\in\mathcal{E}, y\in\mathcal{Y}, \quad \sum_{y\in\mathcal{Y}} \mu_v(y) = 1, v\in\mathcal{V}, \qquad \boldsymbol{\mu}\geq \mathbf{0}$$

Derive the LP dual and show that it can be expressed as an unconstrained problem

$$\boldsymbol{\varphi}^* = \operatorname*{argmin}_{\boldsymbol{\varphi}} \left[\sum_{v\in\mathcal{V}} \max_{y\in\mathcal{Y}} q_v^{\boldsymbol{\varphi}}(x_v, y) + \sum_{\{v,v'\}\in\mathcal{E}} \max_{(y,y')\in\mathcal{Y}^2} g_{vv'}^{\boldsymbol{\varphi}}(y,y') \right]$$

where

$$\begin{aligned} g_{vv'}^{\boldsymbol{\varphi}}(y,y') &= g_{vv'}(y,y') + \varphi_{vv'}(y) + \varphi_{v'v}(y'), & \{v,v'\}\in\mathcal{E}, y,y'\in\mathcal{Y} \\ q_v^{\boldsymbol{\varphi}}(y) &= q_v(y) - \sum_{v'\in\mathcal{N}(v)} \varphi_{vv'}(y), & v\in\mathcal{V}, y\in\mathcal{Y} \end{aligned}$$

Assignment 5. Consider a max-sum classifier for playing Sudoku:

$$\boldsymbol{y}^* = h(\boldsymbol{x}; \boldsymbol{q}, \boldsymbol{g}) = \operatorname*{argmax}_{\boldsymbol{y}\in\mathcal{Y}^{\mathcal{V}}} \left(\sum_{v\in\mathcal{V}} q(x_v, y_v) + \sum_{\{v,v'\}\in\mathcal{E}} g(y_v, y_{v'}) \right)$$

where

- $\mathcal{V} = \{(i,j)\in\mathbb{N}^2 \mid 1\leq i\leq 9, 1\leq j\leq 9\}$
- $\mathcal{E} = \{\{(i,j),(i',j')\} \mid i=i' \vee j=j' \vee (\lceil i/3\rceil = \lceil i'/3\rceil \wedge \lceil j/3\rceil = \lceil j'/3\rceil)\}$
- $\boldsymbol{x} = (x_v \in \{\square, 1, \ldots, 9\} \mid v\in\mathcal{V})$
- $\boldsymbol{y} = (y_v \in \{1, \ldots, 9\} \mid v\in\mathcal{V})$
- $q\colon \{\square, 1, \ldots, 9\} \times \{1, \ldots, 9\} \to \mathbb{R}$
- $g\colon \{1, \ldots, 9\}^2 \to \mathbb{R}$

Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an example of the Sudoku assignment and its correct solution, respectively. Derive a variant of the Perceptron algorithm which finds the quality functions $\boldsymbol{q}$ and $\boldsymbol{g}$ such that $\hat{\boldsymbol{y}} = h(\hat{\boldsymbol{x}}; \boldsymbol{q}, \boldsymbol{g})$ and the max-sum problem $\mathcal{P} = (\mathcal{V}, \mathcal{E}, \boldsymbol{q}, \boldsymbol{g}, \hat{\boldsymbol{x}})$ has a strictly trivial equivalent. Apply the algorithm on a particular example of the Sudoku puzzle and try to interpret the learned quality functions $(\boldsymbol{q}, \boldsymbol{g})$.