STRUCTURED MODEL LEARNING (SS2015)
6.SEMINAR

Assignment 1 Let $\{(\boldsymbol{x}^1, y^1), \ldots, (\boldsymbol{x}^m, y^m)\} \in (\mathbb{R}^n \times \{+1, -1\})^m$ be a set of linearly separable training examples. Let $\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{w}^*, \boldsymbol{x} \rangle = 0\}$ be an optimal separating hyperplane where

$$\boldsymbol{w}^* \in \operatorname*{argmax}_{\boldsymbol{w} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}} \min_{i \in \{1,\ldots,m\}} \frac{y^i \langle \boldsymbol{w}, \boldsymbol{x}^i \rangle}{\|\boldsymbol{w}\|} .$$

Show that $\boldsymbol{w}^* = \lambda \boldsymbol{u}^*$ for some $\lambda > 0$ where

$$\boldsymbol{u}^* = \operatorname*{argmin}_{\boldsymbol{u} \in \mathbb{R}^n} \frac{1}{2} \|\boldsymbol{u}\|^2 \qquad \text{s.t.} \qquad y^i \langle \boldsymbol{u}, \boldsymbol{x}^i \rangle \geq 1, \quad i \in \{1, \ldots, m\}.$$

Assignment 2 Consider learning of a linear classifier $h(\boldsymbol{x}; \boldsymbol{w}) = \operatorname{argmax}_{\boldsymbol{y} \in \mathcal{Y}} \langle \boldsymbol{w}, \boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{y}) \rangle$ by regularized risk minimization

$$\boldsymbol{w}^* = \operatorname*{argmin}_{\boldsymbol{w} \in \mathbb{R}^n} \left[\frac{\lambda}{2} \|\boldsymbol{w}\|^2 + \frac{1}{m} \sum_{i=1}^m \hat{\ell}(\boldsymbol{x}^i, \boldsymbol{y}^i, \boldsymbol{w}) \right] \tag{1}$$

where

$$\hat{\ell}(\boldsymbol{x}^i, \boldsymbol{y}^i, \boldsymbol{w}) = \max_{\boldsymbol{y} \in \mathcal{Y}} \left[\ell(\boldsymbol{y}^i, \boldsymbol{y}) + \langle \boldsymbol{w}, \boldsymbol{\Psi}(\boldsymbol{x}^i, \boldsymbol{y}) \rangle - \langle \boldsymbol{w}, \boldsymbol{\Psi}(\boldsymbol{x}^i, \boldsymbol{y}^i) \rangle \right]$$

is a margin-rescaling proxy of a loss $\ell \colon \mathcal{Y}^2 \to \mathbb{R}_+$ such that $\ell(\boldsymbol{y}, \boldsymbol{y}') = 0$ iff $\boldsymbol{y} = \boldsymbol{y}'$. Reformulate the problem (1) as a quadratic program and derive its dual.

Assignment 3 Let $f(\boldsymbol{u}, \boldsymbol{v})$ be a function which is convex both in $(\boldsymbol{u}, \boldsymbol{v})$ and let $\mathcal{V}$ be a convex nonempty set. Show that the function

$$g(\boldsymbol{u}) = \inf_{\boldsymbol{v} \in \mathcal{V}} f(\boldsymbol{u}, \boldsymbol{v})$$

is convex in $\boldsymbol{u}$ provided $g(\boldsymbol{u}) > -\infty$.

Assignment 4 We have shown on the lecture that the Lagrange dual of the reduced problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^n} \left[\lambda \Omega(\boldsymbol{w}) + \max_{i \in \mathcal{I}} \left(b_i + \langle \boldsymbol{w}, \boldsymbol{a}_i \rangle \right) \right]$$

reads

$$\max_{\boldsymbol{\alpha} \in \mathcal{U}} \left[-\lambda \Omega^*(-\lambda^{-1} \mathbf{A} \boldsymbol{\alpha}) + \langle \boldsymbol{\alpha}, \boldsymbol{b} \rangle \right]$$

where $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^m \mid \|\boldsymbol{u}\|_1 = 1, \boldsymbol{u} \geq \boldsymbol{0}\}$ denotes m-dimensional simplex and $\mathcal{I} = \{1, \ldots, m\}$. Show that the dual can be derived differently, namely by applying the Fenchel duality and proving that

$$f(\boldsymbol{v}) = \max_{i \in \mathcal{I}} \left[v_i + b_i \right]$$

and

$$f^*(\boldsymbol{\alpha}) = \begin{cases} -\langle \boldsymbol{\alpha}, \boldsymbol{b} \rangle & \text{if} \quad \boldsymbol{\alpha} \in \mathcal{U} \\ +\infty & \text{otherwise} \end{cases}$$

are Fenchel conjugates.

Hint: prove that f is the Fenchel conjugate of f^*.

Assignment 5 Consider a minimization problem

$$\min_{\boldsymbol{w}\in\mathbb{R}^n} F_t(\boldsymbol{w}) \quad \text{s.t.} \quad \langle \boldsymbol{c}_i, \boldsymbol{w}\rangle \leq d_i \,, i \in \{1,\ldots,p\} \tag{2}$$

where

$$F_t(\boldsymbol{w}) = \max_{i\in\{1,\ldots,t\}} \Big[\langle \boldsymbol{g}_i, \boldsymbol{w} - \boldsymbol{w}_i\rangle + F(\boldsymbol{w}_i)\Big]$$

is a cutting plane model of some convex function F. Convert the problem (2) to a linear program and derive its dual.

Assignment 6 Consider problem of learning a linear SVM classifier

$$h(\boldsymbol{x}; \boldsymbol{w}, w_0) = \text{sign}(\langle \boldsymbol{w}, \boldsymbol{x}\rangle + w0)$$

from a training set $\{(\boldsymbol{x}^1, y^1), \ldots, (\boldsymbol{x}^m, y^m)\} \in (\mathbb{R}^n \times \{+1,-1\})^m$ by solving

$$(\boldsymbol{w}^*, w_0^*) = \operatorname*{argmin}_{\boldsymbol{w}\in\mathbb{R}^n, w_0\in\mathbb{R}} \Big[\frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{m}\sum_{i=1}^m \xi_i\Big] \quad \text{s.t.} \quad \left\{ \begin{array}{rcll} y^i(\langle \boldsymbol{w}, \boldsymbol{x}^i\rangle + w_0) & \geq & 1 - \xi_i\,, & i \in \{1,\ldots,m\} \\ \xi_i & \geq & 0\,, & i \in \{1,\ldots,m\} \end{array} \right. \tag{3}$$

Note that the bias w_0 is not contained in the quadratic regularizer. Convert the problem (3) to an unconstrained convex problem

$$\boldsymbol{w}^* = \operatorname*{argmin}_{\boldsymbol{w}\in\mathbb{R}^n} \Big[\frac{\lambda}{2}\|\boldsymbol{w}\|^2 + R(\boldsymbol{w})\Big]$$

and derive an algorithm for evaluating $R(\boldsymbol{w})$ and $R'(\boldsymbol{w})$.