Combinatorial Optimization in Computer Vision
Lecture: F. R. Schmidt and C. Domokos
Exercises: T. Möllenhoff and T. Windheuser
Winter Semester 2015/2016

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 12

Room: 02.09.023

Tuesday, 02.02.2016, 14:15-15:45

Submission deadline: Tuesday, 02.02.2016, 11:15 , Room 02.09.023

1 Parameter Learning (15 Points)

Exercise 1 (2 Points)**.** Let $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ be a set of given observations. By assuming that w is a random vector with prior distribution $p(w)$, show that the following posterior distribution can be written as:

$$p(w|\mathcal{D}) = p(w) \prod_{n=1}^{N} \frac{p(y^n|x^n, w)}{p(y^n|x^n)}.$$

Exercise 2 (2 Points)**.** Calculate the expected loss $\mathbb{E}_{y\sim d(y|x)}\left[\Delta_H(y, f(x))\right]$ of the Hamming loss:

$$\Delta_H(y, y') = \frac{1}{|V|} \sum_{i\in V} [\![y_i \neq y'_i]\!],$$

where $d(y|x)$ denotes the true data distribution.

Exercise 3 (6 Points)**.** Consider the negative regularized conditional log-likelihood (cf. Lecture 22):

$$\mathcal{L}(w) = \lambda \|w\|^2 + \sum_{n=1}^{N} \langle w, \varphi(x^n, y^n)\rangle + \sum_{n=1}^{N} \log Z(x^n, w). \tag{1}$$

It has been shown in the lecture that the gradient of (1) w.r.t. w is given as

$$\nabla_w \mathcal{L}(w) = 2\lambda w + \sum_{n=1}^{N} \left[\varphi(x^n, y^n) - \mathbb{E}_{y\sim p(y|x^n)}[\varphi(x^n, y)]\right].$$

Show that the Hessian of (1) is given as

$$H_w \mathcal{L}(w) = 2\lambda I + \sum_{n=1}^{N} \mathbb{E}_{y\sim p(y|x^n)}[\varphi(x^n, y)\varphi(x^n, y)^\mathsf{T}] - \mathbb{E}_{y\sim p(y|x^n)}[\varphi(x^n, y)]\mathbb{E}_{y\sim p(y|x^n)}[\varphi(x^n, y)]^\mathsf{T},$$

and argue that it is positive definite for $\lambda > 0$.

Exercise 4 (2 Points)**.** Let $\xi \sim \mathcal{U}(-1, 1)$ be a continuous random variable which follows a uniform distribution on the interval $[-1, 1]$. Calculate the cumulative distribution function of the variable $\nu \sim \xi^2$.

Exercise 5 (3 Points)**.** Compute the *subdifferential* at a point $x \in \mathbb{R}^n$

$$\partial f(x) = \{w \in \mathbb{R}^n \mid f(x) + \langle w, y - x \rangle \leq f(y), \forall y \in \mathbb{R}^n\},$$

of the following convex functions $f : \mathbb{R}^n \to \mathbb{R}$:

a) $f(x) = \langle c, x \rangle$, where $c \in \mathbb{R}^n$ is a constant

b) $f(x) = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$

c) $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$