

## Weekly Exercise 11

Dr. Csaba Domokos and Lingni Ma  
 Technische Universität München, Computer Vision Group  
 June 28, 2016 (submission deadline: July 05, 2016)

### Parameter Learning

(8 Points)

**Exercise 1** (2 Points). Let  $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$  be a given set of identically and independently distributed (*i.i.d.*) observations. By assuming that  $\mathbf{w}$  is a random vector with prior distribution  $p(\mathbf{w})$ , show that the following *posterior distribution*  $p(\mathbf{w}|\mathcal{D})$  can be written as:

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

**Exercise 2** (6 Points). Consider the *negative regularized conditional log-likelihood*:

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

It has been shown in the lecture that the gradient of  $\mathcal{L}(\mathbf{w})$  w.r.t.  $w$  is given as

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

Show that the Hessian of  $\mathcal{L}(\mathbf{w})$  is given as

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

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

### Programming

(8 Points)

**Exercise 3 (Gibbs sampling, 8 Points)**. Let us consider the problem of binary image segmentation and let us solve it by performing *probabilistic inference* via Gibbs sampling. In this particular exercise, we are going to design a *cow-detector* for the images in Figure 1, which should label a pixel as foreground if it belongs to a cow, and background otherwise.

We define the following energy function for  $\mathbf{y} \in \{0, 1\}^{\mathcal{V}}$  such that 0 denotes the background and 1 denotes the foreground:

$$E(\mathbf{y}) = \sum_{i \in \mathcal{V}} E_i(y_i) + w \sum_{(i,j) \in \mathcal{E}} E_{ij}(y_i, y_j) ,$$

where  $w \in \mathbb{R}^+$  is a parameter, and  $\mathcal{V}$  stands for the set of pixels and  $\mathcal{E}$  includes all pairs of 4-neighboring pixels.

To define the unary energy functions  $E_i$ , use the provided \*.yml files. Each test image has its own data file, specified by the same filename. In each data file, you can read out a  $H \times W$  array of float numbers. The  $H$  and  $W$  are the image height and width, and each float value  $p_i$  corresponds to the probability of that the given pixel belongs to the foreground. We provide the `cow_detector.cpp` to demonstrate how to load a data file and read out the corresponding probability values. The unary energy functions  $E_i$  for all  $i \in \mathcal{V}$  are then defined as,

$$\begin{aligned} E_i(y_i = 0) &= -\log(1 - p_i) \\ E_i(y_i = 1) &= -\log(p_i) . \end{aligned}$$

The pairwise energy is defined as the *contrast-sensitive Potts-model* for all  $(i, j) \in \mathcal{E}$ ,

$$E_{ij}(y_i, y_j; x_i, x_j) = \exp(-\lambda \|x_i - x_j\|^2) \mathbb{I}[y_i \neq y_j] .$$

where  $\lambda = 0.5$ .

Implement the Gibbs sampling algorithm to achieve *probabilistic inference* and calculate the *binary segmentation* as well. Choose different values for  $w$  and compare your results. Give the range of  $w$  that generate the best segmentation performance.



Figure 1: The test images for binary image segmentation to detect cows.