HW Assignment 5, Theory

Problems marked with a ($*$) are mandatory for ITCS 8156 students. Bonus problems are optional, solving them will result in extra points.

1 Gradient of Logistic Regression (20 points)

Prove that the gradient (with respect to $\mathbf{w}$) of the negative log-likelihood error function for binary logistic regression corresponds to the formula shown in the lecture:

$$\nabla_{\mathbf{w}} E(\mathbf{w}) = \sum_{n=1}^{N} (h_n - t_n)\mathbf{x}_n \tag{1}$$

2 Binary Softmax Regression (20 points)

Show that Logistic Regression is a special case of Softmax Regression. That is to say, if $\mathbf{w_1}$ and $\mathbf{w_2}$ are the parameter vectors of a Softmax Regression model for the case of two classes, then there exists a parameter vector $\mathbf{w}$ for Logistic Regression that results in the same classification as the Softmax Regression model, for any example $\mathbf{x}$.

3 Logistic Regression in `sklearn` ($*$) (20 points)

In some ML packages, the objective function for binary logistic regression is formulated to appear similar with the SVM setting, where the labels $t_n \in \{-1, +1\}$ (instead of $\{0, +1\}$), and where a hyper-parameter C expresses the trade-off between model complexity and training error:

$$E(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C * \sum_{n=1}^{N} \ln(e^{-t_n(\mathbf{w}^T\mathbf{x}_n)} + 1) \tag{2}$$

- Show that the sum in the second term is equal with the negative log-likelihood.
- Compute the C parameter such that minimizing this objective is equivalent (has the same solution) as minimizing the standard objective shown on the slides in which the regularization parameter α or λ is multiplied with the L2 norm term.

4 Gradient of Softmax Regression (20 bonus points)

Prove that the gradient (with respect to $\mathbf{w}_k$) of the negative log-likelihood error function for regularized softmax regression corresponds to the formula shown in the lecture, for each class $k \in [1..K]$:

$$\nabla_{\mathbf{w}_k} E(\mathbf{w}) = -\frac{1}{N} \sum_{n=1}^{N} (\delta_k(t_n) - p(C_k|\mathbf{x}_n))\mathbf{x}_n + \alpha\mathbf{w}_k \tag{3}$$

5 Submission

Submit your responses on Canvas as one file named `theory.pdf`. It is recommended to use an editor such as Latex or Word or Jupyter-Notebook that allows editing and proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it on Canvas. Make sure that your writing is legible and the scan has good quality.