Theoretical task 3: logistic regression

Due Tuesday, January 27 (seminar).

1. Let $\sigma(z)$ define the logistic function: $\sigma(z) = \frac{1}{1+e^{-z}}$. Show that $\sigma'(z) = \sigma(z)(1 - \sigma(z))$

2. Suppose $y_i \in \{0, 1\}$ is the class of i-th observation and the first feature is constant: $x_i^1 = 1\,\forall i$. Logistic regression weight parameter w is found by minimization of negative log-likelihood:

$$NLL(w) = -\sum_{i=1}^{N} \ln[\sigma(w^T x_i)^{\mathbb{I}[y_i=1]} \times \left(1 - \sigma(w^T x_i)\right)^{\mathbb{I}[y_i=0]}] = -\sum_{i=1}^{N} \left[y_i \ln \sigma(w^T x_i) + (1 - y_i) \ln(1 - \sigma(w^T x_i))\right]$$

   Please find the gradient $g = \frac{d}{dw} NLL(w)$ and the Hessian matrix $H = \frac{d}{dw} g(w)^T$. Prove that the Hessian matrix will be positive definite (assuming $X = [x_1, x_2, ...x_N] \in \mathbb{R}^{D \times N}$ is full rank).

3. Suppose we optimize L_2 regularized logistic regression with cost function: $J(w) = NLL(w) + \lambda||w||_2^2$. Please write whether each of the following statements is true or not?

   (a) $J(w)$ has multiple locally optimum solutions.

   (b) Let $\widehat{w} = \arg\min_w J(w|\text{training set})$ be the global minimum. Then $\widehat{w}$ is sparse, meaning that it has many zero entries.

   (c) Let training data be linearly separable. Then during optimization some weight w_j may become infinite when $\lambda = 0$.

   (d) $NLL(\widehat{w})$ evaluated on training set will always increase with the increase of λ.

   (e) $NLL(\widehat{w})$ evaluated on testing set will always increase with the increase of λ.