Logistic Regression

Dr François Pitié

Problem 1. Recall that for a feature vector $\mathbf{x}$ and parameter vector $\mathbf{w}$, the likelihood of success is modelled as:

$$h_{\mathbf{w}}(\mathbf{x}) = p(y=1|\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}^\top \mathbf{x}}}$$

Show that the likelihood of failure is $p(y=0|\mathbf{x}) = h_{\mathbf{w}}(-\mathbf{x})$.

Problem 2. Recall that

$$E(\mathbf{w}) = \sum_{i=1}^{n} -y_i \log(h_{\mathbf{w}}(\mathbf{x}_i)) - (1-y_i) \log(1-h_{\mathbf{w}}(\mathbf{x}_i))$$

$$\text{with} \quad h_{\mathbf{w}}(\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}^\top \mathbf{x}}}$$

Show that

$$\frac{\partial E}{\partial \mathbf{w}} = \sum_{i=1}^{n} (h_{\mathbf{w}}(\mathbf{x}_i) - y_i)\, \mathbf{x}_i$$

Problem 3. Adapt the cross entropy loss function below to include a regularisation term (ie. penalty on $\mathbf{w}$ for deviating from the null vector):

$$E(\mathbf{w}) = \sum_{i=1}^{n} -y_i \log(h_{\mathbf{w}}(\mathbf{x}_i)) - (1-y_i) \log(1-h_{\mathbf{w}}(\mathbf{x}_i))$$

Derive the gradient $\frac{\partial E'}{\partial \mathbf{w}}$ for the new expression of the loss function $E'(\mathbf{w})$.

Problem 4. Compute the gradient $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$ of

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + b \qquad \text{, with } \mathbf{A} \text{ symmetric} \tag{1}$$

$$f(\mathbf{x}) = \cos(\mathbf{a}^\top \mathbf{x}) \tag{2}$$

$$f(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \exp\left(-\frac{\|\mathbf{x}-\mathbf{a}_i\|^2}{2}\right) \tag{3}$$