Logistic Regression

Dr François Pitié

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

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

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

Problem 2. Recall that

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

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

Show that

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

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

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

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

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

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

$$[2] \quad f(\mathsf{x}) = \cos(\mathsf{a}^\top \mathsf{x}) \tag{2}$$

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