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 $p(y=0|\mathsf{x}) = h_{\mathsf{w}}(-\mathsf{x})$.

answer:

$$p(y=0|\mathsf{x}) = 1 - p(y=1|\mathsf{x}) = 1 - \frac{1}{1+e^{-\mathsf{w}^\top \mathsf{x}}} \tag{1}$$

$$= \frac{1+e^{-\mathsf{w}^\top \mathsf{x}} - 1}{1+e^{-\mathsf{w}^\top \mathsf{x}}} \times \frac{e^{+\mathsf{w}^\top \mathsf{x}}}{e^{+\mathsf{w}^\top \mathsf{x}}} \tag{2}$$

$$= \frac{1}{e^{+\mathsf{w}^\top \mathsf{x}}+1} = h_{\mathsf{w}}(-\mathsf{x}) \tag{3}$$

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

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

answer:

$$\frac{\partial}{\partial \mathsf{w}} h_{\mathsf{w}}(\mathsf{x}) = \frac{\mathsf{x} e^{-\mathsf{w}^\top \mathsf{x}}}{\left(1 + e^{-\mathsf{x}^\top}\right)^2} = h_{\mathsf{w}}(\mathsf{x})(1 - h_{\mathsf{w}}(\mathsf{x}))\mathsf{x} \tag{4}$$

$$\frac{\partial}{\partial \mathsf{w}} \log(h_{\mathsf{w}}(\mathsf{x})) = \frac{1}{h_{\mathsf{w}}(\mathsf{x})} \frac{\partial}{\partial \mathsf{w}} h_{\mathsf{w}}(\mathsf{x}) = (1 - h_{\mathsf{w}}(\mathsf{x}))\mathsf{x} \tag{5}$$

$$\frac{\partial}{\partial \mathsf{w}} \log(1 - h_{\mathsf{w}}(\mathsf{x})) = \frac{1}{1 - h_{\mathsf{w}}(\mathsf{x})} \frac{\partial}{\partial \mathsf{w}} (1 - h_{\mathsf{w}}(\mathsf{x})) = -h_{\mathsf{w}}(\mathsf{x})\mathsf{x} \tag{6}$$

Say $E(\mathsf{w}) = \sum_{i=1}^{n} e_i(\mathsf{w})$. Then we have

$$\frac{\partial}{\partial \mathsf{w}} e_i(\mathsf{w}) = \begin{cases} -\frac{\partial}{\partial \mathsf{w}} \log(h_{\mathsf{w}}(\mathsf{x}_i)) = -(1 - h_{\mathsf{w}}(\mathsf{x}_i))\mathsf{x}_i = (h_{\mathsf{w}}(\mathsf{x}_i) - y_i)\mathsf{x}_i & \text{if } y_i = 1 \\ -\frac{\partial}{\partial \mathsf{w}} \log(1 - h_{\mathsf{w}}(\mathsf{x}_i)) = h_{\mathsf{w}}(\mathsf{x}_i)\mathsf{x}_i = (h_{\mathsf{w}}(\mathsf{x}_i) - y_i)\mathsf{x}_i & \text{if } y_i = 0 \end{cases}$$

Thus

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

Problem 3. Adapt the cross entropy loss function below to include a regularisation term (ie. penalty on $\mathbf{w}$ 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 $E'(\mathbf{w})$.

answer:

There is nothing complicated here. For instance a L2 regularisation would look like this:

$$E'(\mathbf{w}) = E(\mathbf{w}) + \lambda \|\mathbf{w}\|^2$$

with $\lambda > 0$ (recall that $\|\mathbf{w}\|^2 = \mathbf{w}^\top \mathbf{w}$).

The gradient is also quite simple:

$$\frac{\partial E'}{\partial \mathbf{w}} = \frac{\partial E}{\partial \mathbf{w}} + 2\lambda \mathbf{w} \tag{7}$$

$$= \sum_{i=1}^{n} (h_{\mathbf{w}}(\mathbf{x}_i) - y_i)\, \mathbf{x}_i + 2\lambda \mathbf{w} \tag{8}$$

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

$$[1] \quad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + b \qquad \text{, with } \mathbf{A} \text{ symmetric} \tag{9}$$

$$[2] \quad f(\mathbf{x}) = \cos(\mathbf{a}^\top \mathbf{x}) \tag{10}$$

$$[3] \quad f(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{2}\right) \tag{11}$$

answer:

$$[1] \quad \frac{\partial f}{\partial \mathbf{x}} = \mathbf{A}\mathbf{x} \tag{12}$$

$$[2] \quad \frac{\partial f}{\partial \mathbf{x}} = -\sin(\mathbf{a}^\top \mathbf{x})\mathbf{a} \tag{13}$$

$$[3] \quad \frac{\partial f}{\partial \mathbf{x}} = -\sum_{i=1}^{n} \lambda_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{2}\right)(\mathbf{x} - \mathbf{a}_i) \tag{14}$$