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

answer:

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

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

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

Problem 2. Recall that

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

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

Show that

$$\frac{\partial E}{\partial \mathbf{W}} = \sum_{i=1}^{n} \left(h_{\mathbf{W}}(\mathbf{X}_i) - y_i \right) \mathbf{X}_i$$

answer:

$$\frac{\partial}{\partial \mathbf{W}} h_{\mathbf{W}}(\mathbf{X}) = \frac{\mathbf{X} e^{-\mathbf{W}^{\top} \mathbf{X}}}{\left(1 + e^{-\mathbf{X}^{\top}}\right)^{2}} = h_{\mathbf{W}}(\mathbf{X})(1 - h_{\mathbf{W}}(\mathbf{X}))\mathbf{X}$$
(4)

$$\frac{\partial}{\partial \mathbf{W}} \log(h_{\mathbf{W}}(\mathbf{X})) = \frac{1}{h_{\mathbf{W}}(\mathbf{X})} \frac{\partial}{\partial \mathbf{W}} h_{\mathbf{W}}(\mathbf{X}) = (1 - h_{\mathbf{W}}(\mathbf{X}))\mathbf{X}$$
 (5)

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

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

$$\frac{\partial}{\partial \mathbf{W}} e_i(\mathbf{W}) = \begin{cases} -\frac{\partial}{\partial \mathbf{W}} \log(h_{\mathbf{W}}(\mathbf{X}_i)) = -(1 - h_{\mathbf{W}}(\mathbf{X}_i)) \mathbf{X}_i = (h_{\mathbf{W}}(\mathbf{X}_i) - y_i) \mathbf{X}_i & \text{if } y_i = 1 \\ -\frac{\partial}{\partial \mathbf{W}} \log(1 - h_{\mathbf{W}}(\mathbf{X}_i)) = h_{\mathbf{W}}(\mathbf{X}_i) \mathbf{X}_i = (h_{\mathbf{W}}(\mathbf{X}_i) - y_i) \mathbf{X}_i & \text{if } y_i = 0 \end{cases}$$

Thus

$$\frac{\partial E}{\partial \mathbf{W}} = \sum_{i=1}^{n} \left(h_{\mathbf{W}}(\mathbf{X}_i) - y_i \right) \mathbf{X}_i$$

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

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

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:

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

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

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

[1]
$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} + b$$
 , with **A** symmetric (9)

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

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

answer:

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

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

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