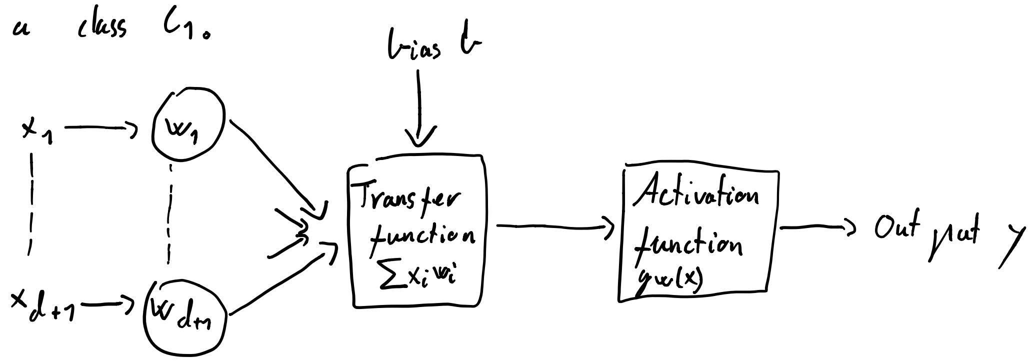
1 Logistic Regression .

Logistic regression is a tool for binary classification that "uses" one neuron to determin the probability that input x belong in a class C_1.



We have : $P(x \in C_1 | x) = g_w(x) = \dfrac{1}{1 + e^{-w^T x}}$

$$P(x \in C_2 | x) = 1 - P(x \in C_1 | x) = 1 - g_w(x)$$

$$g_w(x) \in [0,1]$$

Loss function $E(w) = -\dfrac{1}{N} \displaystyle\sum_{n=1}^{N} t^n \ln(y^n) + (1 - t^n) \ln(1 - t^n)$

measures how well our hypothesis function $g_w(x) = y^n$ does on N data points. t^n = target value for example n

Goal: $t_n = y^n$ for all n.

* Softmax regression is a generalization of logistic regression to a multi-class classification.

1.1

Show that $-\dfrac{\partial E^n(w)}{\partial w_j} = (t^n - y^n)\, x_j^n$

We have $E^n(w) = -t^n \ln(g_w^n) + (1-t^n)\ln(1-t^n)$

$$\Longrightarrow \quad \frac{\partial E^n(w)}{\partial w_j} = -t^n \frac{1}{g_w^n} \frac{\partial g_w^n}{\partial w_j} \qquad \text{Applies chain rule.}$$

Using the hint : $$\frac{\partial E^n(w)}{\partial w_j} = -t^n \frac{1}{\cancel{g_w^n}}\, x_j^n\, \cancel{g_w^n}\,(1-g_w^n)$$

$$-\frac{\partial E^n(w)}{\partial w_j} = t^n x_j^n (1-g_w^n) = (t^n - y^n)\, x_j^n$$

1.2

We have for the Soft max regression cost function

$$E = -\sum_{n=1}^{N}\sum_{K'=1}^{C} t_{K'}^n \ln(y_{K'}^n) + t_K^n \ln(y_K^n) \quad , \quad K' \neq K$$

where : $y_K^n = \dfrac{e^{a_K^n}}{\sum_{K'} e^{a_{K'}^n}}$ is the Soft max function with

$a_K^n = w_K^T x^n$ called net input to y_K and

y_K^n is the probability that x is a member of class K. Note $\sum^{K} y_K^n = 1$

We have two cases : $K \neq K'$ and $K = K'$,

hence we get :

Forsøk 3:

$$E^n(w) = -\overbrace{\sum_{k'=1}^{c} t_{k'}^n \ln\left(y_{k'}^n\right)}^{\text{Case } k' \neq k} + \overbrace{t_k^n \ln\left(y_k^n\right)}^{\text{Case } k' = k}$$

$$= -\sum_{k'=1}^{c} t_{k'}^n \left(\ln\left(e^{w_{k'}^T x^n}\right) - \ln\left(\sum_{k'} e^{w_{k'}^T x^n}\right)\right) + t_k^n \ln\left(\frac{e^{w_k^T x^n}}{\sum_{k''} e^{w_{k''}^T x^n}}\right)$$

$$\frac{\partial E^n(w)}{\partial w_{kj}} = -\sum_{k=1}^{c} t_{k'}^n \left(\cancel{\left(\frac{1}{e^{w_k^T x^n}} \cdot e^{w_k^T x^n} \cdot x_j^n \right)} - \left(\frac{1}{\sum e^{w_{k'}^T x^n}} e^{w_k^T x^n} x_j^n\right)\right) +$$

$$t_k^n \left(\left(\frac{1}{\cancel{e^{w_k^T x^n}}} \cdot \cancel{e^{w_k^T x^n}} \cdot x_j^n \right) - \left(\frac{1}{\sum e^{w_{k'}^T x^n}} e^{w_k^T x^n} x_j^n\right)\right)$$

$$\frac{\partial E^n(w)}{\partial w_{kj}} = -\sum_{k'=1}^{c} t_{k'}^n \left(x_j^n y_k^n\right) + t_k^n x_j^n \left(1 - y_k^n\right)$$

$$-\frac{\partial E^n(w)}{\partial w_{kj}} = x_j^n y_k^n \sum_{k'=1}^{c} t_{k'}^n + t_k^n x_j^n - x_j^n y_k^n$$

$$= x_j^n y_k^n \left(\cancel{\sum t_{k'}^n} - 1\right) + t_k^n x_j^n$$

$$= \underline{\underline{x_j^n \left(t_k^n - y_k^n\right)}}$$

Task 2.2

$$C(w) = \|w_{i,j}\|^2 = \sum_{i,j} w_{i,j}^2$$

$$\frac{\partial C(w)}{\partial w_{i,j}} = 2 w_{i,j}$$

Gradient descent with regularization then becomes:

$$dw = \frac{dE(w)}{dw} + \lambda \frac{dC(w)}{dw}$$