# 2. HW

Montag, 8. April 2024 17:39

## Aufgabenstellung:

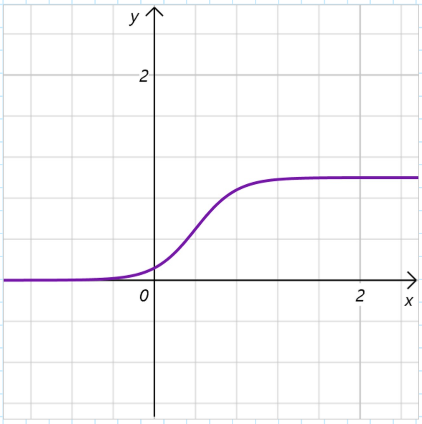
Klassifizierung

## Sigmoid Funktion:

$$y(a) = \frac{1}{1+e^{-a}}$$

$a \rightarrow \vec{w}^T \vec{x} \in \mathbb{R}$

$$y(\vec{x}) = \frac{1}{1+e^{-\vec{w}^T\vec{x}}}$$




## Lernen der Parameter:

### Binomialverteilung:

$$p(t_n \mid \vec{w}) = y_n^{t_n} (1-y_n)^{1-t_n}$$

### Likelihood Funktion:

$$p(\vec{t} \mid \vec{w}) = \prod_{n=1}^{N} y_n^{t_n} (1-y_n)^{1-t_n}$$

$t_n$ ... Sollwert aus Datensatz

$y_n$ ... Output mit aktuellem $\vec{w}$

## Schrittweise Optimierung von $\vec{w}$

1.) Fehler berechnen

2.) Gradient des Fehlers $\nabla E$ analysieren

3.) Gewichtung updaten

4.3.3 (Bishop, 2006)

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (\mathbf{w}^{\mathrm{T}} \phi_n - t_n)\phi_n = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{t}$$

$$\mathbf{H} = \nabla\nabla E(\mathbf{w}) = \sum_{n=1}^{N} \phi_n \phi_n^{\mathrm{T}} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$$

$$\begin{aligned}\mathbf{w}^{(\mathrm{new})} &= \mathbf{w}^{(\mathrm{old})} - (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\left\{\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w}^{(\mathrm{old})} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{t}\right\} \\ &= (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}\end{aligned}$$

## Algorithmus:

1.) Inputs:

1.) Inputs:

Init: Anzahl an Spalten vom Trainingsdatensatz für weigths

Train: PCA projezierten Features (Matrix), binäre Labels, Anzahl an trainings Iterationen

2.) Train:

weights

$w_{new} = w_{old} - H^{-1} \nabla E(\omega)$ ... für $\sigma(z) = \frac{1}{1+e^{-z}}$, wobei: $z = w^T x$

$H = \Phi^T R \Phi$

$R = I \cdot y_n \cdot (1 - y_n)$

$y_n = \sigma(w^T x_n) = \frac{1}{1+e^{-w^T x}}$ ... Varianzen der binomial Verteilung

$\Phi$ ... Messdaten wobei die erste Spalte mit 1en für den BIAS hinzugefügt werden sollte

$\nabla E(w) = \Phi^T (y_n - t)$ ... wobei y Label Matrix ist