Lineare Modell

$$\vec{y} = \vartheta_1 \cdot \vec{x} + \vartheta_0$$

Mean-Squared-Error (MSE)

$$MSE(\vartheta_0, \vartheta_1) = \frac{1}{n-1} \sum_{i=1}^{n} (\vartheta_1 \cdot x_i + \vartheta_0 - y_i)^2$$

Um MSE möglichst zu verringern, suchen wir den Tiefpunkt der Funktion, indem wir die Ableitung gleich null setzen.

Da die Funktion auf zwei Inputs basiert, müssen wir zwei Ableitungen bilden, einmal nach ϑ_0 und einmal nach ϑ_1:

$$\frac{\partial MSE}{\partial \vartheta_0} = \frac{1}{n-1} \sum_{i=1}^{n} 2(\vartheta_1 \cdot x_i + \vartheta_0 - y_i) = 0$$

$$\iff 2 \sum_{i=1}^{n} (\vartheta_1 \cdot x_i + \vartheta_0 - y_i) = 0$$

$$\iff \sum_{i=1}^{n} (\vartheta_1 \cdot x_i + \vartheta_0 - y_i) = 0$$

$$\iff \sum_{i=1}^{n} \vartheta_1 x_i + \sum_{i=1}^{n} \vartheta_0 - \sum_{i=1}^{n} y_i = 0$$

$$\iff \vartheta_1 \sum_{i=1}^{n} x_i + n \cdot \vartheta_0 - \sum_{i=1}^{n} y_i = 0$$

$$\iff \vartheta_1 \cdot n \cdot \bar{x} + n \cdot \vartheta_0 - n \cdot \bar{y} = 0$$

$$\iff \vartheta_1 \cdot \bar{x} + \vartheta_0 - \bar{y} = 0$$

$$\iff \vartheta_0 = \bar{y} - \vartheta_1 \cdot \bar{x}$$

$$\frac{\partial MSE}{\partial \vartheta_1} = \frac{1}{n-1} \sum_{i=1}^{n} 2x_i(\vartheta_1 \cdot x_i + \vartheta_0 - y_i) = 0$$

$$\iff \frac{1}{n-1} \sum_{i=1}^{n} 2x_i(\vartheta_1 \cdot x_i + \bar{y} - \vartheta_1 \cdot \bar{x} - y_i) = 0$$

$$\iff 2 \sum_{i=1}^{n} x_i(\vartheta_1 \cdot x_i + \bar{y} - \vartheta_1 \cdot \bar{x} - y_i) = 0$$

$$\iff \sum_{i=1}^{n} x_i(\vartheta_1 \cdot x_i + \bar{y} - \vartheta_1 \cdot \bar{x} - y_i) = 0$$

$$\iff \sum_{i=1}^{n} (x_i \cdot \vartheta_1 \cdot x_i + x_i \cdot \bar{y} - x_i \cdot \vartheta_1 \cdot \bar{x} - x_i \cdot y_i) = 0$$

$$\iff \sum_{i=1}^{n} x_i \cdot \vartheta_1 \cdot x_i + \sum_{i=1}^{n} x_i \cdot \bar{y} - \sum_{i=1}^{n} x_i \cdot \vartheta_1 \cdot \bar{x} - \sum_{i=1}^{n} x_i \cdot y_i = 0$$

$$\iff \sum_{i=1}^{n} x_i^2 \cdot \vartheta_1 + \bar{y} \cdot n \cdot \bar{x} - \bar{x} \cdot n \cdot \vartheta_1 \cdot \bar{x} - \sum_{i=1}^{n} x_i \cdot y_i = 0$$

$$\iff \vartheta_1 \sum_{i=1}^{n} x_i^2 + \bar{y} \cdot n \cdot \bar{x} - \bar{x} \cdot n \cdot \vartheta_1 \cdot \bar{x} - \sum_{i=1}^{n} x_i \cdot y_i = 0$$

$$\iff \vartheta_1 \sum_{i=1}^{n} x_i^2 - \bar{x} \cdot n \cdot \vartheta_1 \cdot \bar{x} = \sum_{i=1}^{n} x_i \cdot y_i - \bar{y} \cdot n \cdot \bar{x}$$

$$\iff \vartheta_1 (\sum_{i=1}^{n} x_i^2 - \bar{x} \cdot n \cdot \bar{x}) = \sum_{i=1}^{n} x_i \cdot y_i - \bar{y} \cdot n \cdot \bar{x}$$

$$\iff \vartheta_1 = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \bar{y} \cdot n \cdot \bar{x}}{\sum_{i=1}^{n} x_i^2 - \bar{x} \cdot n \cdot \bar{x}}$$

$$\iff \vartheta_1 = \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y}}{\sum_{i=1}^{n} x_i x_i - \sum_{i=1}^{n} x_i \bar{x}}$$

$$\iff \vartheta_1 = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$$

Ergebnis: $\vartheta_0 = \bar{y} - \vartheta_1 \bar{x}$ und $\vartheta_1 = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$

Alternative Schreibweise des Ergebnisses: $\vartheta_0 = \bar{y} - \vartheta_1 \bar{x}$ und $\vartheta_1 = r_{x,y} \frac{s_y}{s_x}$

- Varianz: $Var_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Standardabweichung: $s_x = \sqrt{Var_x}$
- Kovarianz: $Cov_{x,y} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- Pearson-Korrelation: $r_{x,y} = \frac{Cov_{x,y}}{s_x s_y}$

Interessantes Merkmal: $MSE(\bar{y}, 0) = Var_y$

Unerklärte Varianz: $\frac{MSE(\vartheta_0, \vartheta_1)}{Var_y}$

Erklärte Varianz: $R^2 = 1 - \frac{MSE(\vartheta_0, \vartheta_1)}{Var_y}$

Allgemeine Lineare Regression

Hier haben wir nun für jeden Ausgabewert eine Liste von Eingabewerten.

Das Modell ist dann eine Funktion $F(\vec{x}) \approx y$.

Im Folgenden steht X für die Matrix von Eingaben und $\vec{y}$ steht für die Liste von Ausgaben.
Für jede Eingabeliste $\vec{x}^{(i)}$ (der Länge p) haben wir einen Ausgabewert $y^{(i)}$

Wir definieren unser Modell:

$$F(\vec{x}) = b + \sum_{j=1}^{p} \vec{w}_j \cdot \vec{x}_j$$

MSE

Def.:

$$MSE = \frac{1}{n-1} \sum_{i=1}^{n} (F(\vec{x}^{(i)}) - y^{(i)})^2 = \frac{1}{n-1} (X\vec{w} - \vec{y})^T \cdot (X\vec{w} - \vec{y})$$

Hilfsregeln für Ableitung des MSEs nachher

$$\frac{\partial}{\partial \vec{x}_j} \vec{x}^T C \vec{x} = \frac{\partial}{\partial \vec{x}_j} \vec{x}^T (\sum_{i=1}^{n} C_{k,i} \cdot \vec{x}_i)$$

$$= \frac{\partial}{\partial \vec{x}_j} \sum_{k=1}^{n} \sum_{i=1}^{n} x_k C_{k,i} x_i$$

$$= \frac{\partial}{\partial \vec{x}_j} \sum_{k=1}^{n} \sum_{i=1}^{n} C_{k,i} x_k x_i$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} C_{k,i} \cdot \frac{\partial}{\partial \vec{x}_j} (x_k x_i)$$

Bevor wir das vereinfachen können, müssen wir das Kronecker Delta definieren: $\delta_{i,j} = 1$ wenn $i = j$ gilt und $\delta_(i, j) = 0$ andernfalls

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} C_{k,i} \cdot (\delta_{j,k} \vec{x}_i + \delta_{j,i} \vec{x}_k)$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} C_{k,i} \delta_{j,k} \vec{x}_i + \sum_{k=1}^{n} \sum_{i=1}^{n} C_{k,i} \delta_{j,i} \vec{x}_k$$

$$= \sum_{i=1}^{n} C_{j,i} \vec{x}_i + \sum_{k=1}^{n} C_{k,j} \vec{x}_k$$

$$= (C\vec{x})_j + (C^T \vec{x})_j$$

$$= (C + C^T) \vec{x}$$

$$= 2C\vec{x}$$

Jetzt kommt noch die zweite Hilfsregel, die wir uns beweisen müssen:

$$\frac{\partial}{\partial \vec{x}_j} \vec{b}^T A \vec{x} = \frac{\partial}{\partial \vec{x}_j} \vec{b}^T (\sum_{i=1}^{n} A_{k,i} \vec{x}_i)$$

$$= \vec{b}^T \cdot \frac{\partial}{\partial \vec{x}_j} (\sum_{i=1}^{n} A_{k,i} \vec{x}_i)$$

$$= \vec{b}^T \cdot (\sum_{i=1}^{n} A_{k,i} \delta_{j,i})$$

$$= A^T \cdot \vec{b}$$

Notiz zur Matrixmultiplikation:

```
| 1 |^T → | 1  3  1  4  3  6  1 |
| 3 |
| 1 |
| 4 |
| 3 |
| 6 |
| 1 |

Matrix-Multiplation:

          | 3 |
| 1 2 | * | 4 | = 1*3 + 2*4

| 1  2 |   | 5  6 |   | 5*1 + 7*2    6*1 + 8*2 |
| 3  4 | * | 7  8 | = | 5*3 + 7*4    6*3 + 8*4 |
```

MSE ableiten & Nullsetzen

$$\frac{\partial MSE}{\partial \vec{w}} = \frac{\partial}{\partial \vec{w}} \cdot \frac{1}{n-1} \cdot (X\vec{w} - \vec{y})^T \cdot (X\vec{w} - \vec{y})$$

$$= \frac{\partial}{\partial \vec{w}} \cdot \frac{1}{n-1} \cdot (\vec{w}^T X^T - \vec{y}^T) \cdot (X\vec{w} - \vec{y})$$

$$= \frac{\partial}{\partial \vec{w}} \cdot \frac{1}{n-1} \cdot (\vec{w}^T X^T X\vec{w} - \vec{w}^T X^T \vec{y} - \vec{y}^T X\vec{w} + \vec{y}^T \vec{y})$$

$$= \frac{1}{n-1} \cdot \frac{\partial}{\partial \vec{w}} (\vec{w}^T X^T X\vec{w} - 2\vec{y}^T X\vec{w} + \vec{y}^T \vec{y})$$

$$= \frac{1}{n-1} \cdot (\frac{\partial}{\partial \vec{w}} \vec{w}^T X^T X\vec{w} - \frac{\partial}{\partial \vec{w}} 2\vec{y}^T X\vec{w} + 0)$$

Die erste Hilfsregel kann für den ersten Summanden angewendet werden, wobei $\vec{w} \to \vec{x}$ und $X^T X \to C$:

$$= \frac{1}{n-1} \cdot (2X^T X\vec{w} - \frac{\partial}{\partial \vec{w}} 2\vec{y}^T X\vec{w})$$

Die zweite Hilfsregel kann für den zweiten Summanden angewendet werden, wobei $\vec{y} \to \vec{b}, X \to A, \vec{w} \to \vec{x}$:

$$= \frac{1}{n-1} \cdot (2X^T X\vec{w} - 2X^T \vec{y})$$

$$0 = \frac{1}{n-1} \cdot (2X^T X\vec{w} - 2X^T\vec{y})$$

$$0 = 2X^T X\vec{w} - 2X^T\vec{y}$$

$$0 = X^T X\vec{w} - X^T\vec{y}$$

$$X^T X\vec{w} = X^T\vec{y}$$

Das wird auch als Normalgleichung bezeichnet.