Logistic regression:

$$P(y_1 \mid x) = \frac{P(y_1|x)P(x)}{P(y_1)} = \frac{P(y_1|x)P(x)}{P(y_1|x)P(x)+P(y_2|x)P(x)} = \frac{1}{1+\frac{P(y_2|x)P(x)}{P(y_1|x)P(x)}} = \frac{1}{1+exp(-\alpha)}$$

mit

$$\alpha = -ln\left(\frac{P(y_2|x)P(x)}{P(y_1|x)P(x)}\right)$$

Regression: berechne α als $\alpha = w_{k,0} + \sum_{d=1}^{D} w_{k,d}x_i = \sum_{d=0}^{D} w_{k,d}x_i$

Lernen: Minimiere Log Likelihood $-P(y_1 \mid w_0, w_1, ..., w_N)$ $-P(y_1 \mid w_0, w_1, ..., w_N) = -ln\prod_{n=1}^{N}\prod_{k=1}^{K} P(y_k \mid x_n)^{y_{nk}} = -\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk}lnP(y_k \mid x_k)$

Nur 2 Klassen: $-P(y_1 \mid w_0, w_1, ..., w_N)$ $-P(y_1 \mid w_0, w_1, ..., w_N) = -ln\prod_{n=1}^{N} P(y_1 \mid x_n)^{y_{n1}}P(y_2 \mid x_n)^{y_{2k}} = -\sum_{n=1}^{N} y_{1k}lnP(y_{1k} \mid x_k) + y_{2k}lnP(y_{2k} \mid x_k)$

Einsetzen der oberen Definition:
$\mathcal{L} = -P(y_1 \mid w_0, w_1, ..., w_N) = -\sum_{n=1}^{N}\sum_{k=1}^{D} y_{nd}ln\frac{1}{1+exp(\sum_{i=0}^{D} w_{nd}x_d)}$

Ableitung der Log Likelihood:
$\frac{\partial\mathcal{L}}{\partial w_{ml}} = -\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk}(1+exp(\sum_{d=0}^{D} w_{nd}x_d))(-x_l exp(-\sum_{d=0}^{D} w_{nd}x_d))\frac{1}{(1+exp(\sum_{i=0}^{D} w_{nd}x_d))^2}$

das ist gleich:

$-\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk}(-x_l exp(\sum_{d=0}^{D} w_{nd}x_d))\frac{1}{(1+exp(\sum_{i=0}^{D} w_{nd}x_d))}$

das ist gleich:

$\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk}\frac{x_l}{(1+exp(-\sum_{i=0}^{D} w_{nd}x_d))}$

$P(y_1 \mid x) = \frac{1}{1+\frac{P(y_2|x)P(x)}{P(y_1|x)P(x)}}$

$\mathcal{L}(w) = \prod_{i=1}^{N} P(y_i \mid x_i, w)$

$$
\begin{aligned}
\mathcal{L}(w) &= \prod_{i=1}^{N} P(y_i \mid x_i, w) \\
&= \prod_{i=1}^{N} P(y_i \mid x_i, w)^{y_i} + (1 - P(y_i \mid x_i, w))^{1-y_i}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\ell(w) &= ln(\mathcal{L}(w)) \\
&= \sum_{i=1}^{N} y_i P(y_i \mid x_i, w) + (1 - y_i)(1 - P(y_i \mid x_i, w))
\end{aligned} \tag{2}
$$

$P(y_1 \mid x_1, x_2) = P(y_1 \mid x)$

$P(y_2 \mid x) = 1 - P(y_1 \mid x)$

Linear regression:
$P(y \mid x, w) = \prod$