logistic regression model

$$P(y=1 \mid x, w) = \psi(w^t x) = \frac{1}{1+\exp(-w^t x)} = \frac{\exp(w^t x)}{1+\exp(w^t x)}$$

Fit the model by the maximum log-likelihood criterion

$$J_L(w) = \sum_{k=1}^{n} \log P(y_k \mid x_k, w)$$

$$= \sum_{k=1}^{n} \left[y_k \log P(y_k=1 \mid x_k, w) + (1-y_k) \log \left(1 - P(y_k=1 \mid x_k, w)\right) \right]$$

$$= \sum_{k=1}^{n} \left[y_k \left(w^t x_k - \log \left(1+\exp(w^t x_k)\right)\right) \right.$$

$$\left. + (1-y_k) \log \frac{1}{1+\exp(w^t x_k)} \right]$$

$$= \sum_{k=1}^{n} y_k w^t x_k - \log \left(1+\exp(w^t x_k)\right)$$

Thus, we set the derivatives of $J_L(w)$ w.r.t. the parameters to zero.

$$\frac{\partial J_L(w)}{\partial w_0} = \sum_{k=1}^{n} y_k - \frac{\exp(w^t x_k)}{1+\exp(w^t x_k)} = \sum_{k=1}^{n} y_k - P(y_k=1 \mid x_k, w) = 0$$

$$\frac{\partial J_L(w)}{\partial w_j} = \sum_{k=1}^{n} y_k x_{kj} - \frac{\exp(w^t x)\, x_{kj}}{1+\exp(w^t x_k)} = \sum_{k=1}^{n} \left(y_k - P(y_k=1 \mid x_k, w)\right) x_{kj} = 0$$

errors:

$$e_k = \left(y_k - P(y_k=1 \mid x_k, w)\right), \quad k=1, \cdots, n$$

stochastic gradient ascent ($J_L(w)$ is concave)

$$w \leftarrow w + \eta \cdot \frac{\partial}{\partial w} J_L^{(k)}(w)$$

$$= w + \eta \cdot \left(y_k - P(y_k=1 \mid x_k, w)\right) x_k$$

η is the learning rate.