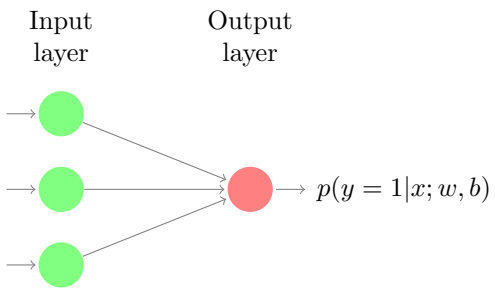
Logistic Regression

Jacob Goldberger



Binary Case

logistic regression model: $\hat{y} = p(y=1|x;w,b) = \frac{1}{1+\exp(-wx-b)}$

Input data: $x_1, ..., x_n \in R^d$
input labels: $y_1, ..., y_n \in \{0, 1\}$

Likelihood: $L(w,b) = \sum_t \log p(y_t|x_t;w,b)$

Gradient ascent algorithm:

$$w \leftarrow w + \epsilon \sum_t (y_t - \hat{y}_t) x_t$$

$$b \leftarrow b + \epsilon \sum_t (y_t - \hat{y}_t)$$

Multi-class Case

logistic regression model: $\hat{y}_i = p(y=i|x;w,b) = \frac{\exp(w_i x + b_i)}{\sum_{j=1}^{k} \exp(w_j x + b_j)}$

Input data: $x_1, ..., x_n \in R^d$
input labels: $y_1, ..., y_n \in \{1, ..., k\}$

Likelihood: $L(w,b) = \sum_t \log p(y_t|x_t;w,b)$

Gradient ascent algorithm:

$$w_i \leftarrow w_i + \epsilon \sum_t (1_{\{y_t=i\}} - \hat{y}_{ti}) x_t$$

$$b_i \leftarrow b_i + \epsilon \sum_t (1_{\{y_t=i\}} - \hat{y}_{ti})$$