1 Die logistische Regression

Ein weiteres Modell zur binären Klassifikation ist die logistische Regression. Die logistische Regression arbeitet mit dem sogenannten Chancenverhältnis t mit der Erfolgswahrscheinlickeit p.

$$t = \frac{p}{1-p} \tag{1}$$

Logarithmiert man dieses Verhältnis, erhällt man nun die logit-Funktion. Diese Funktion hat die schöne Eigenschaft bei 0.5 eine Nullstelle zu haben, während sie darunter negativ und darüber positiv ist, allerdings nie 1 oder 0 erreicht. Diese Funktion setzen wir mit dem Skalarprodukt der Gewichte des Modells und de Eingabe gleich, da wir eine Funktion der Wahrscheinlichkeit haben wollen, die von den jeweiligen Gewichten und den Eingaben ab.

$$\omega \cdot x = \log_{(\frac{p}{1-p})} \tag{2}$$

Umgeformt ergibt dies nach p

$$p(\omega \cdot x) = \frac{e^{\omega \cdot x}}{1 + e^{\omega \cdot x}} \tag{3}$$

Diese Wahrscheinlichkeitsfunktion wird nun also als Aktivierungsfunktion genutzt und es wird diesmal ein Schwellenwert als Wahrscheinlichkeit definiert. Die Sprungfunktion wird nun zur Klassifizierung so definiert.

$$\phi(z) = \begin{Bmatrix} 1, & p(\omega \cdot x) \geq 0.5 \\ 0, & p(\omega \cdot x) \leq 0.5 \end{Bmatrix} \tag{4}$$

Um nun die Gewichtsänderung an die Wahrscheinlichkeit einer Vorhersage anzupassen, wird die Wahrscheinlichkeit, unter der Annahme die Ereignisse sind stochastisch unabhänging, folgendermaßen definiert. Da die einzelnen Ereignisse unabhängig voneinander sind, tritt es bei einem Datensatz der Größe n auf, dass es entweder richtig oder falsch klassifiziert wird Mit j^i als der richtigen Klassifizierung gilt:

$$l(w) = \prod_{i=1}^{n} (p(\omega \cdot x^i))^{(y)^j} (1 - p(\omega \cdot x^i))^{1-(y)^j} \tag{5}$$

Ist also die Vorherage korrekt, es sei j^i also 1, dann fällt der zweite Term weg. Ist Sie es nicht, dann fällt der zweite Term weg und es ergibt sich einfach de Formel für die Erolgswahrscheinlichkeit mit der Anzahl der richtigen Klassifizierungen im Datensatz multipliziert mit der Wahrscheinlichkeit der

Klassifizierung als nicht-zugehörig mit der Anzahl dieser Klassifizierungen im Exponenten. Da der natürliche Logarithmus dieser Funktion einfacher zu maximieren ist, gilt:

$$\ln(l(w)) = \sum_{i=1}^{n} \ln((p(\omega \cdot x^i))^{(y)^j} (1 - p(\omega \cdot x^i))^{1-(y)^j}) \tag{6}$$

$$\ln(l(w)) = \sum_{i=1}^{n} \ln((p(\omega \cdot x^i))^{(y)^j}) + \ln((1 - p(\omega \cdot x^i))^{1-(y)^j}) \tag{7}$$

$$\ln(l(w)) = \sum_{i=1}^{n} (y)^j \ln(p(\omega \cdot x^i)) + (1 - (y)^j) \ln(1 - p(\omega \cdot x^i)) \tag{8}$$

Nun definieren wir diese Funktion als Zielfunktion und verwenden das Gradient-Descent-Verfahren des Adaline-Modells. Damit berechnet sich ein einzelner Term zu:

$$\Delta\omega = \eta \frac{\partial}{\partial \omega^i} (y)^j \ln(p(\omega \cdot x^i)) + (1 - (y)^j) \ln(1 - p(\omega \cdot x^i)) \tag{9}$$

$$\Delta\omega = -\eta \frac{\partial}{\partial \omega^i} (y)^j \ln(\frac{e^{\omega \cdot x}}{1 + e^{\omega \cdot x}}) + (1 - (y)^j) \ln(1 - \frac{e^{\omega \cdot x}}{1 + e^{\omega \cdot x}}) \tag{10}$$

$$\Delta\omega = \eta \frac{\partial}{\partial \omega^i} (y)^j (\omega \cdot x - \ln(1 + e^{\omega \cdot x}) + (1 - (y)^j) \ln(\frac{1 + e^{\omega \cdot x} - e^{\omega \cdot x}}{1 + e^{\omega \cdot x}}) \tag{11}$$

$$\Delta\omega = -\eta \frac{\partial}{\partial \omega^i} (y)^j (\omega \cdot x - \ln(1 + e^{\omega \cdot x}) + (1 - (y)^j) \ln(\frac{1}{1 + e^{\omega \cdot x}}) \tag{12}$$

$$\Delta\omega = \eta \frac{\partial}{\partial \omega^i} (y)^j (\omega \cdot x - \ln(1 + e^{\omega \cdot x}) + (1 - (y)^j)(- \ln(1 + e^{\omega \cdot x})) \tag{13}$$

$$\Delta\omega = \eta \frac{\partial}{\partial \omega^i} ((y)^j \omega \cdot x - \ln(1 + e^{\omega \cdot x})) \tag{14}$$

$$\Delta\omega = \eta(x^i (y)^j - \frac{1}{1 + e^{\omega \cdot x}} e^{\omega \cdot x} x^i) \tag{15}$$

$$\Delta\omega = \eta x^i (j^i - p(\omega \cdot x)) \tag{16}$$

Dies ist der gleiche Korrekturterm wie im Adaline-Modell.