

# Logistic Regression

Michael Staniek

University of Heidelberg

November 26, 2017

# Inhalt

## 1 Blockvektoren

# Recap von letzter Woche

- Wir haben bisher 1 Gewichtsvektor, der + und - unterscheiden kann
- Wie kriegen wir es hin, mehrere Klassen zu unterscheiden?
- Mehrere Gewichtsvektoren!
- $w_1=(0.1, 1)$ ,  $w_2=(10, -2)$
- Welcher Gewichtsvektor für ein Trainingsbeispiel den höchsten Score liefert, gewinnt.
- $x=(1, 1)$
- $w_1*x=1.1$
- $w_2*x=8$
- Klasse von  $w_2$  wird genommen.
- Sollte aber  $w_1$  sein, was tun?
- $w_1=w_1+x$
- $w_2=w_2-x$

# Als Blockvektoren

- $(w_{1.1}, w_{1.2}, w_{2.1}, w_{2.2})$
- $(0.1, 1, 10, -2)$
- $x=(1, 1)$
- $\phi(x, c1) = (1, 1, 0, 0)$
- $\phi(x, c2) = (0, 0, 1, 1)$
- $w * \phi(x, c1) = 1.1$
- $w * \phi(x, c2) = 8$
- Klasse von  $w_2$  wird genommen.
- Sollte aber  $w_1$  sein, was tun?
- $w = w + \phi(x, c1)$
- $w = w - \phi(x, c2)$
- oder:  $w = w + \phi(x, c1) - \phi(x, c2)$

# Als Blockvektoren (Logistic Regression)

- $(w_{1.1}, w_{1.2}, w_{2.1}, w_{2.2})$
- $(0.1, 1, 10, -2)$
- $x = (1, 1)$
- $\phi(x, c_1) = (1, 1, 0, 0)$
- $\phi(x, c_2) = (0, 0, 1, 1)$
- $p(c_1|x) = \frac{e^{\phi(x, c_1) * w}}{\sum_{c \in C} e^{\phi(x, c) * w}}$
- $e^{\phi(x, c_1) * w} = e^{1.1} = 3.00$
- $e^{\phi(x, c_2) * w} = e^8 = 2980.95$
- $p(c_1|x) = \frac{e^{1.1}}{e^{1.1} + e^8} = 0.001$
- $p(c_2|x) = \frac{e^8}{e^{1.1} + e^8} = 0.9989$

# Updateregel (Logistic Regression)

- $(w_{1.1}, w_{1.2}, w_{2.1}, w_{2.2})$
- $w = (0.1, 1, 10, -2)$
- $x = (1, 1)$
- $\phi(x, c_1) = (1, 1, 0, 0)$
- $\phi(x, c_2) = (0, 0, 1, 1)$
- $p(c_1|x) = \frac{e^{1.1}}{e^{1.1} + e^8} = 0.001$
- $p(c_2|x) = \frac{e^8}{e^{1.1} + e^8} = 0.9989$
- Wie vorhin, erste Klasse ist aber richtig :
- $w\text{-grad} = \phi(xt, ct) - \sum_{c \in C} p(c|xt) * \phi(xt, c)$
- $w\text{-grad} = \phi(xt, c_1) - (p(c_1|xt) * \phi(xt, c_1) + p(c_2|xt) * \phi(xt, c_2))$
- $w\text{-grad} = (1, 1, 0, 0) - (0.001 * (1, 1, 0, 0) + 0.9989 * (0, 0, 1, 1))$
- $w\text{-grad} = (1, 1, 0, 0) - ((0.001, 0.001, 0, 0) + (0, 0, 0.9989, 0.9989))$
- $w\text{-grad} = (1, 1, 0, 0) - (0.001, 0.001, 0.9989, 0.9989)$
- $w\text{-grad} = (0.999, 0.999, -0.9989, -0.9989)$

# Updateregel (Logistic Regression)

- $w\text{-gradient}=(0.999, 0.999, -0.9989, -0.9989)$
- $w=(0.1, 1, 10, -2)$
- neues  $w$ :  $w=(0.1, 1, 10, -2)+0.1*(0.999, 0.999, -0.9989, -0.9989)$
- neues  $w$ :  $w=(0.1, 1, 10, -2)+0.1*(0.999, 0.999, -0.9989, -0.9989)$
- neues  $w$ :  $(0.1999, 1.0999, 9.90011, -2.09989)$
- $\phi(x, c1) = (1, 1, 0, 0)$
- $\phi(x, c2) = (0, 0, 1, 1)$
- $e^{\phi(x,c1)*w} = e^{1.2998} = 3.66$
- $e^{\phi(x,c1)*w} = e^{7.8} = 2440.6$
- $p(c1|x) = \frac{e^{1.2998}}{e^{1.2998}+e^{7.8}} = 0.0015$
- $p(c2|x) = \frac{e^{7.8}}{e^{1.2998}+e^{7.8}} = 0.9985$

Jetzt habt ihr in der Übung bereits Vektoren aufaddiert, das heisst ab jetzt könnt ihr Numpy auch benutzen