Логистическая регрессия

November 12, 2022

Краткая теоретическая справка

Пусть вероятности двух событий связаны соотношениями:

$$P\left\{y=1|x\right\}=\sigma\left(z\right),$$

$$P\left\{y=0|x\right\}=1-\sigma\left(z\right),$$

$$z=b+\left(\vec{x},\vec{\theta}\right),$$

$$\sigma\left(z\left(x\right)\right)=\frac{1}{1+e^{-z(x)}},$$

$$\frac{\partial}{\partial x}\sigma\left(z\left(x\right)\right)=\frac{\partial z\left(x\right)}{\partial x}\left(1-\sigma\left(z\left(x\right)\right)\right)\sigma\left(z\left(x\right)\right).$$

Тогда:

$$P\left\{y=1|x\right\}=\sigma\left(b+\left(\vec{x},\vec{\theta}\right)\right),$$

$$P\left\{y=0|x\right\}=1-\sigma\left(b+\left(\vec{x},\vec{\theta}\right)\right).$$

В общем виде:

$$P\left\{y|x\right\}=\sigma\left(b+\left(\vec{x},\vec{\theta}\right)\right)^{y}\left(1-\sigma\left(b+\left(\vec{x},\vec{\theta}\right)\right)\right)^{1-y}.$$

Максимальное правдоподобие:

$$\left\{b,\vec{\theta}\right\}=argmax\prod_{i}P\left\{y_i|x_i\right\}=argmax\prod_{i}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)^{y_i}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)^{1-y_i}.$$

Максимум этой фукнции аналогичен максимому логарифма этой функции:

$$\left\{b,\vec{\theta}\right\}=argmax\prod_{i}P\left\{y_i|x_i\right\}=argmax\sum_{i}ln\left(P\left\{y_i|x_i\right\}\right),$$

$$argmax\sum_{i}ln\left(P\left\{y_i|x_i\right\}\right)=\sum_{i}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)^{y_i}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)^{1-y_i}\right)=$$

$$\sum_{i}y_i ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)+\left(1-y_i\right)ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right).$$

Для градиентного спуска:

$$\begin{cases}\frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b} = & \frac{\partial\sum_i y_i ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)+(1-y_i)ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)}{\partial b}\\ \frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b\theta_j} = & \frac{\partial\sum_i y_i ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)+(1-y_i)ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)}{\partial b\theta_j}\end{cases}.$$

Распишем все необходимые производные. Для этого сперва випишем соотношения для производных функций от $\sigma\left(\right)$:

$$\frac{\partial}{\partial x}\sigma\left(z\left(x\right)\right)=\frac{\partial z\left(x\right)}{\partial x}\left(1-\sigma\left(z\left(x\right)\right)\right)\sigma\left(z\left(x\right)\right),$$

$$\frac{\partial}{\partial b}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)=\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right),$$

$$\frac{\partial}{\partial \theta_j}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)=x_i^{(j)}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right).$$

Теперь подставим эти соотношения в $\frac{\partial}{\partial b}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)$ и $\frac{\partial}{\partial b}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)$:

$$\frac{\partial}{\partial b}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=\frac{\frac{\partial}{\partial b}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=\frac{\frac{\partial}{\partial b}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right),$$

$$\frac{\partial}{\partial \theta_j}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=\frac{\frac{\partial}{\partial \theta_j}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=\vec{x}_i^{(j)}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right),$$

$$\frac{\partial}{\partial b}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=\frac{\frac{\partial}{\partial b}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)}{1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=$$

$$=-\frac{\frac{\partial}{\partial b}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=-\frac{\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right),$$

$$\frac{\partial}{\partial \theta_j}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=-\frac{\frac{\partial}{\partial \theta_j}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}{1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)}=-\vec{x}_i^{(j)}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right).$$

Окончательно:

$$\begin{cases}\frac{\partial}{\partial b}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)=\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial}{\partial \theta_j}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)=x_i^{(j)}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial}{\partial b}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial}{\partial \theta_j}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=x_i^{(j)}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\\ \frac{\partial}{\partial b}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial}{\partial \theta_j}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)=-x_i^{(j)}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\end{cases}.$$

Градиент примет вид:

$$\begin{cases}\frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b}= & \sum_i y_i\frac{\partial}{\partial b}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)+(1-y_i)\frac{\partial}{\partial b}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\\ \frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b\theta_j}= & \sum_i y_i\frac{\partial}{\partial \theta_j}ln\left(\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)+(1-y_i)\frac{\partial}{\partial \theta_j}ln\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\end{cases},$$

$$\begin{cases}\frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b}= & \sum_i y_i\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)-(1-y_i)\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b\theta_j}= & \sum_i y_i\left(x_i^{(j)}\left(1-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\right)-(1-y_i)x_i^{(j)}\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\end{cases},$$

$$\begin{cases}\frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b}= & \sum_i y_i-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\\ \frac{\partial\sum_i ln(P\{y_i|x_i\})}{\partial b\theta_j}= & \sum_i x_i^{(j)}\left(y_i-\sigma\left(b+\left(\vec{x}_i,\vec{\theta}\right)\right)\right)\end{cases}.$$

Можно переобозначить:

$$b+\left(\vec{x}_i,\vec{\theta}\right)=\left([1,\vec{x}_i]^T,\left[b,\vec{\theta}\right]\right)=\left(\vec{x}_i,\vec{\theta}\right).$$

Получим:

$$\frac{\partial\sum_i ln\left(P\left\{y_i|x_i\right\}\right)}{\partial b\theta_j}=\sum_i x_i^{(j)}\left(y_i-\sigma\left(\left(\vec{x}_i,\vec{\theta}\right)\right)\right),$$

что может быть хаписано в матричном виде:

$$\frac{\partial \sum_i ln\left(P\left\{y_i|x_i\right\}\right)}{\partial b\theta_j} = \sum_i x_i^{(j)} \left(y_i - \sigma\left(\left(\vec{x}_i, \vec{\theta}\right)\right)\right) = X^T \left(\vec{y} - \sigma\left(X\vec{\theta}\right)\right).$$

Итерационная последовательность для логистической регрессии в матричной форме:

$$\vec{\theta}_i = \vec{\theta}_{i-1} + \boldsymbol{X}^T \left(\sigma\left(\boldsymbol{X}\vec{\theta}_{i-1}\right) - \vec{y}\right).$$

Задания

1. Рассмотреть пример с линейной разделимостью двух групп. Любой элемент каждой группы характеризуется двумя параметрами.

2. Рассмотреть пример с нелинейной разделимостью двух групп. В качестве разделителя использовать полином второй степени. Любой элемент каждой группы характеризуется двумя параметрами. Если же вам хочется вызова, то полином произвольной степени и произвольным количеством набора параметров элемента группы.

3. Сравнить результаты с Keras.