Backpropagation za logističku regresiju

May 4, 2023

Postavke

Model je logistička regresija:

$$\hat{y} = \sigma(z)$$

$$z = (\sum_i w_i x_i)$$

Postavke

Model je logistička regresija:

$$\hat{y} = \sigma(z)$$

$$z = (\sum_i w_i x_i)$$

Error je SSE:

$$\mathbb{E} = \frac{1}{2}(t - \hat{y})^2$$

Postavke

Model je logistička regresija:

$$\hat{y} = \sigma(z)$$

$$z = (\sum_i w_i x_i)$$

Error je SSE:

$$\mathbb{E} = \frac{1}{2}(t - \hat{y})^2$$

Aktivacija je sigmoidna:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Postavke

Model je logistička regresija:

$$\hat{y} = \sigma(z)$$

$$z = (\sum_i w_i x_i)$$

Error je SSE:

$$\mathbb{E} = \frac{1}{2}(t - \hat{y})^2$$

Aktivacija je sigmoidna:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Metoda optimizacije je stohastički gradijentni spust:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i}$$

Osnovna pravila deriviranja

1. LD:
$$[f(x)a + g(x)b]' = f(x)'a + g(x)'b$$
2. REC:
$$[\frac{1}{f(x)}]' = -\frac{f'(x)}{f(x)^2}$$
3. CONST:
$$c' = 0$$
4. VAR:
$$\frac{dy}{dz}z = 1$$
5. EXP:
$$[f(x)^n]' = n \cdot f(x)^{n-1} \cdot f'(x)$$
6. CEXP:
$$[e^{f(x)}]' = e^{f(x)} \cdot f'(x)$$
7. CHAIN:
$$\frac{dx}{dz} = \frac{dx}{dy}\frac{dy}{dz}$$

Weight update za logističku regresiju

Za puni GD:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i + \sum_n \eta x_i^{(n)} \hat{y}^{(n)} (1 - \hat{y}^{(n)})(t^{(n)} - \hat{y}^{(n)})$$

Weight update za logističku regresiju

Za puni GD:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i + \sum_n \eta x_i^{(n)} \hat{y}^{(n)} (1 - \hat{y}^{(n)})(t^{(n)} - \hat{y}^{(n)})$$

Mi radimo online SGD:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i + \eta x_i \hat{y}(1 - \hat{y})(t - \hat{y})$$

Weight update za logističku regresiju

Za puni GD:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i + \sum_n \eta x_i^{(n)} \hat{y}^{(n)} (1 - \hat{y}^{(n)})(t^{(n)} - \hat{y}^{(n)})$$

Mi radimo online SGD:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i + \eta x_i \hat{y}(1 - \hat{y})(t - \hat{y})$$

No zašto $\frac{\partial \mathbb{E}}{\partial w_i} = x_i \hat{y}(1 - \hat{y})(t - \hat{y})$?

Za početak, CHAIN...

Prvo s CHAIN raskomadamo:

$$\frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial \hat{y}}{\partial w_i} \frac{d\mathbb{E}}{d\hat{y}}$$

Za početak, CHAIN...

Prvo s CHAIN raskomadamo:

$$\frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial \hat{y}}{\partial w_i} \frac{d\mathbb{E}}{d\hat{y}}$$

Na desni dio ćemo se još vratiti, za sad gledamo $\frac{\partial \hat{y}}{\partial w_i}$. Ponovno primjenimo CHAIN i dobijemo

$$\frac{\partial \hat{y}}{\partial w_i} = \frac{d\hat{y}}{dz} \frac{\partial z}{\partial w_i}$$

Za početak, CHAIN...

Prvo s CHAIN raskomadamo:

$$\frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial \hat{y}}{\partial w_i} \frac{d\mathbb{E}}{d\hat{y}}$$

Na desni dio ćemo se još vratiti, za sad gledamo $\frac{\partial \hat{y}}{\partial w_i}$. Ponovno primjenimo CHAIN i dobijemo

$$\frac{\partial \hat{y}}{\partial w_i} = \frac{d\hat{y}}{dz} \frac{\partial z}{\partial w_i}$$

Kako $z = \sum_i w_i x_i$, $\frac{\partial z}{\partial w_i} = x_i$ (analogno vrijedi i za $\frac{\partial z}{\partial x_i} = w_i$)

Za početak, CHAIN...

Prvo s CHAIN raskomadamo:

$$\frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial \hat{y}}{\partial w_i}\frac{d\mathbb{E}}{d\hat{y}}$$

Na desni dio ćemo se još vratiti, za sad gledamo $\frac{\partial \hat{y}}{\partial w_i}$. Ponovno primjenimo CHAIN i dobijemo

$$\frac{\partial \hat{y}}{\partial w_i} = \frac{d\hat{y}}{dz}\frac{\partial z}{\partial w_i}$$

Kako $z = \sum_i w_i x_i$, $\frac{\partial z}{\partial w_i} = x_i$ (analogno vrijedi i za $\frac{\partial z}{\partial x_i} = w_i$)
Sve što još trebamo pokazati je $\frac{d\hat{y}}{dz} = \hat{y}(1 - \hat{y})$

$$\frac{d\hat{y}}{dz}\frac{1}{1+e^{-z}} = -\frac{\frac{d\hat{y}}{dz}(1+e^{-z})}{(1+e^{-z})^2} \qquad (REC)$$

$$= -\frac{\frac{d\hat{y}}{dz}1 + \frac{d\hat{y}}{dz}e^{-z}}{(1+e^{-z})^2} \qquad (LD)$$

$$= -\frac{0 + \frac{d\hat{y}}{dz}(e^{-z})}{(1+e^{-z})^2} \qquad (CONST)$$

$$= -\frac{e^{-z}\cdot\frac{d\hat{y}}{dz}(-z)}{(1+e^{-z})^2} \qquad (CEXP)$$

$$= -\frac{e^{-z}\cdot\frac{d\hat{y}}{dz}(z)\cdot(-1)}{(1+e^{-z})^2} \qquad (LD)$$

$$= -\frac{e^{-z}\cdot(-1)}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2} \qquad (VAR)$$

$\frac{d\hat{y}}{dz} = \hat{y}(1 - \hat{y})$

Ajmo ovo faktorizirati na sljedeći način:

$$\frac{d\hat{y}}{dz} = \frac{e^{-z}}{(1 + e^{-z})^2} = \frac{1}{1 + e^{-z}} \cdot \frac{e^{-z}}{1 + e^{-z}}$$

Ajmo ovo faktorizirati na sljedeći način:

$$\frac{d\hat{y}}{dz} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}}$$

Prisjetimo se da $\frac{1}{1+e^{-z}} = \hat{y}$

$\frac{d\hat{y}}{dz} = \hat{y}(1 - \hat{y})$

Ajmo ovo faktorizirati na sljedeći način:

$$\frac{d\hat{y}}{dz} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}}$$

Prisjetimo se da $\frac{1}{1+e^{-z}} = \hat{y}$
Sad treba samo pojednostavniti

$$\frac{e^{-z}}{1+e^{-z}} = \frac{(1+e^{-z})-1}{1+e^{-z}} = \frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}} = 1 - \frac{1}{1+e^{-z}} = 1 - \hat{y}$$

Ajmo ovo faktorizirati na sljedeći način:

$$\frac{d\hat{y}}{dz} = \frac{e^{-z}}{(1 + e^{-z})^2} = \frac{1}{1 + e^{-z}} \cdot \frac{e^{-z}}{1 + e^{-z}}$$

Prisjetimo se da $\frac{1}{1+e^{-z}} = \hat{y}$
Sad treba samo pojednostavniti

$$\frac{e^{-z}}{1 + e^{-z}} = \frac{(1 + e^{-z}) - 1}{1 + e^{-z}} = \frac{1 + e^{-z}}{1 + e^{-z}} - \frac{1}{1 + e^{-z}} = 1 - \frac{1}{1 + e^{-z}} = 1 - \hat{y}$$

Dakle, $\frac{d\hat{y}}{dz} = \hat{y}(1 - \hat{y})$

Ostaje nam jos drugi dio, a to je $\frac{d\mathbb{E}}{d\hat{y}}$.

Ostaje nam jos drugi dio, a to je $\frac{d\mathbb{E}}{d\hat{y}}$.
Prisjetimo se da koristimo SSE, i to u varijanti za stohastički gradijentni spust, $\mathbb{E} = \frac{1}{2}(t - y)^2$ (za običan gradijentni spust ide još suma po svim predikcijama na datasetu).

Ostaje nam jos drugi dio, a to je $\frac{d\mathbb{E}}{d\hat{y}}$.
Prisjetimo se da koristimo SSE, i to u varijanti za stohastički gradijentni spust, $\mathbb{E} = \frac{1}{2}(t - y)^2$ (za običan gradijentni spust ide još suma po svim predikcijama na datasetu).

$$\frac{d\mathbb{E}}{d\hat{y}}[\frac{1}{2}(t - \hat{y})^2] =_{LD} \frac{1}{2}\frac{d\mathbb{E}}{d\hat{y}}(t - \hat{y})^2 =_{EXP} \frac{1}{2} \cdot 2 \cdot (t - \hat{y}) \cdot \frac{d\mathbb{E}}{d\hat{y}}(t - \hat{y})$$

$$=_{LD} (t - \hat{y})[\frac{d\mathbb{E}}{d\hat{y}}t - \frac{d\mathbb{E}}{d\hat{y}}\hat{y}] =_{CONST,VAR} (t - \hat{y}) \cdot (0 - 1) = -1 \cdot (t - \hat{y})$$

Konačni weight update za logističku regresiju prema specifikacijama (SSE, Logistička, SGD)

Prisjetimo se:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} \qquad \frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial z}{\partial w_i} \frac{d\hat{y}}{dz} \frac{d\mathbb{E}}{d\hat{y}}$$

Konačni weight update za logističku regresiju prema specifikacijama (SSE, Logistička, SGD)

Prisjetimo se:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} \qquad \frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial z}{\partial w_i} \frac{d\hat{y}}{dz} \frac{d\mathbb{E}}{d\hat{y}}$$

$$\frac{d\hat{y}}{dz} = \hat{y}(1-\hat{y}) \qquad \frac{d\mathbb{E}}{d\hat{y}} = -(t-\hat{y}) \qquad \frac{\partial z}{\partial w_i} = x_i$$

Konačni weight update za logističku regresiju prema specifikacijama (SSE, Logistička, SGD)

Prisjetimo se:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} \qquad \frac{\partial \mathbb{E}}{\partial w_i} = \frac{\partial z}{\partial w_i} \frac{d\hat{y}}{dz} \frac{d\mathbb{E}}{d\hat{y}}$$

$$\frac{d\hat{y}}{dz} = \hat{y}(1-\hat{y}) \qquad \frac{d\mathbb{E}}{d\hat{y}} = -(t-\hat{y}) \qquad \frac{\partial z}{\partial w_i} = x_i$$

Sve sto nam preostaje jest sastaviti weight update formulu:

$$w_i' = w_i - \eta \frac{\partial \mathbb{E}}{\partial w_i} = w_i - \eta \frac{\partial z}{\partial w_i} \frac{d\hat{y}}{dz} \frac{d\mathbb{E}}{d\hat{y}} = w_i - \eta(x_i \cdot \hat{y}(1-\hat{y}) \cdot -(t-\hat{y})) =$$

$$= w_i + \eta x_i \hat{y}(1-\hat{y})(t-\hat{y})$$