

Logisztikus regresszió: maximum likelihood becslés

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

2020 október 19.

Hipotézis:

$$h(\mathbf{x}; \mathbf{a}) := g(\mathbf{a} \cdot \mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{a} \cdot \mathbf{x})}}$$

Hogyan használjuk a hipotézist?

Annak valószínűsége, hogy $y = 1$ vagy $y = 0$:

$$P(y = 1|x; \mathbf{a}) = g(\mathbf{a} \cdot \mathbf{x})$$

$$P(y = 0|x; \mathbf{a}) = 1 - g(\mathbf{a} \cdot \mathbf{x})$$

Valószínűségi értelmezés

Hipotézis:

$$h(\mathbf{x}; \mathbf{a}) := g(\mathbf{a} \cdot \mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{a} \cdot \mathbf{x})}}$$

Hogyan használjuk a hipotézist?

Annak valószínűsége, hogy $y = 1$ vagy $y = 0$:

$$P(y = 1|x; \mathbf{a}) = g(\mathbf{a} \cdot \mathbf{x})$$

$$P(y = 0|x; \mathbf{a}) = 1 - g(\mathbf{a} \cdot \mathbf{x})$$

Ez kompaktabban:

$$P(y|x; \mathbf{a}) = (g(\mathbf{a} \cdot \mathbf{x}))^y (1 - g(\mathbf{a} \cdot \mathbf{x}))^{1-y}$$

Maximum likelihood becslés¹

Tekintsünk egy $\{(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}) \dots (\mathbf{x}^{(N)}, y^{(N)})\}$ adathalmazt!

- ha az egyes $(\mathbf{x}^{(j)}, y^{(j)})$ adatok függetlenek
- akkor annak valószínűsége, hogy N mérés után a fenti adathalmazt kapjuk:

$$L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) = \prod_{i=1}^N P(y^{(i)} | \mathbf{x}^{(i)}; \mathbf{a}) = \prod_{i=1}^N (g(\mathbf{a} \cdot \mathbf{x}^{(i)}))^{y^{(i)}} (1 - g(\mathbf{a} \cdot \mathbf{x}^{(i)}))^{1-y^{(i)}}$$

¹Andrew Ng, Coursera alapján

Maximum likelihood becslés¹

Tekintsünk egy $\{(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}) \dots (\mathbf{x}^{(N)}, y^{(N)})\}$ adathalmazt!

- ha az egyes $(\mathbf{x}^{(j)}, y^{(j)})$ adatok függetlenek
- akkor annak valószínűsége, hogy N mérés után a fenti adathalmazt kapjuk:

$$L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) = \prod_{i=1}^N P(y^{(i)} | \mathbf{x}^{(i)}; \mathbf{a}) = \prod_{i=1}^N (g(\mathbf{a} \cdot \mathbf{x}^{(i)}))^{y^{(i)}} (1 - g(\mathbf{a} \cdot \mathbf{x}^{(i)}))^{1-y^{(i)}}$$

- az $L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ segítségével egy maximum likelihood becslést adhatunk az \mathbf{a} paramétervektorra
- az \mathbf{a} paramétervektor legvalószínűbb értéke: amelyre $L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ maximális.

¹Andrew Ng, Coursera alapján

Maximum likelihood becslés

Az $L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ helyett tekinthetjük pl. a negatív logaritmusát is:

$$\begin{aligned} J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) &= -\frac{1}{N} \log L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) \\ &= -\frac{1}{N} \sum_{i=1}^N y^{(i)} \log(g(\mathbf{a} \cdot \mathbf{x}^{(i)})) + (1 - y^{(i)}) \log(1 - g(\mathbf{a} \cdot \mathbf{x}^{(i)})) \end{aligned}$$

Maximum likelihood becslés

Az $L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ helyett tekinthetjük pl. a negatív logaritmusát is:

$$\begin{aligned} J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) &= -\frac{1}{N} \log L(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) \\ &= -\frac{1}{N} \sum_{i=1}^N y^{(i)} \log(g(\mathbf{a} \cdot \mathbf{x}^{(i)})) + (1 - y^{(i)}) \log(1 - g(\mathbf{a} \cdot \mathbf{x}^{(i)})) \end{aligned}$$

Mivel a negatív logaritmust vettük, ezért a maximum likelihood becslés:

$$\min_{\mathbf{a}} [J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})]$$

$J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ minimumát kell megtalálni \mathbf{a} függvényében