Lagrangian und Naive Bayes

Michael Staniek

November 23, 2017

1 Lagrangian example

https://www.math.ucdavis.edu/~thomases/W11_16C1_lec_2_4_11.pdf
$f(x,y) = 100 * x^{\frac{3}{4}} * y^{\frac{1}{4}}$
Contraint: $200x + 250y = 50000$
$F(x,y,\lambda) = 100 * x^{\frac{3}{4}} * y^{\frac{1}{4}} - \lambda(200x + 250y - 50000)$
$\Delta xF(x,y,\lambda) = 75 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} - \lambda 200 = 0$
$\Delta yF(x,y,\lambda) = 25 * x^{\frac{3}{4}} * y^{\frac{-3}{4}} - \lambda 250 = 0$
$\Delta \lambda F(x,y,\lambda) = -200x - 250y + 50000 = 0$
Als erstes nach lambda in der ersten Ableitung lösen:
$75 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} - \lambda 200 = 0$
$75 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} = \lambda 200$
$0.375 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} = \lambda$
Dann die zweite Gleichung nehmen:
$\Delta yF(x,y,\lambda) = 25 * x^{\frac{3}{4}} * y^{\frac{-3}{4}} - \lambda 250 = 0$
Dort das Lambda einsetzen:
$= 25 * x^{\frac{3}{4}} * y^{\frac{-3}{4}} - 0.375 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} * 250 = 0$
$25 * x^{\frac{3}{4}} * y^{\frac{-3}{4}} = 0.375 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} * 250$
Vereinfachen:
$25 * x^{\frac{3}{4}} * y^{\frac{-3}{4}} = 93.75 * x^{\frac{-1}{4}} * y^{\frac{1}{4}}$
JEtzt durch $x^{\frac{-1}{4}}$ teilen:
$25 * x * y^{\frac{-3}{4}} = 93.75 * y^{\frac{1}{4}}$
Und jetzt durch $y^{\frac{-3}{4}}$ teilen:
$25 * x = 93.75 * y$
$25 * x = 93.75 * y$
$x = 3.75 * y$
Dann nehmen wir die letzte Formel:
$\Delta \lambda F(x,y,\lambda) = -200x - 250y + 50000 = 0$

Und setzen anstatt x dann halt die 3.75*y ein:
$-200 * 3.75 * y - 250y + 50000 = 0$
$-200 * 3.75 * y - 250y + 50000 = 0$
$-1000y + 50000 = 0$
$50000 = 1000y$
$50 = y$
Das können wir hier einsetzen: $-200x - 250y + 50000 = 0$
$-200x - 12500 + 50000 = 0$
$-200x + 37500 = 0$
$37500 = 200x$
$187.5 = x$

Und schlussendlich hier reintun:
$0.375 * x^{\frac{-1}{4}} * y^{\frac{1}{4}} = \lambda$
$0.375 * 187.5^{\frac{-1}{4}} * 50^{\frac{1}{4}} = \lambda$
$0.26947 = \lambda$

2 Naive Bayes

$$P = \underset{P(y)}{argmax} \prod_{t=1}^{T} (P(y_t) \prod_{i=1}^{M} P(\phi_i(x_t)|y_t))$$
$$P = \underset{P(y)}{argmax} \sum_{t=1}^{T} (\log P(y_t) + \sum_{i=1}^{M} \log P(\phi_i(x_t)|y_t))$$
$$P = \underset{P(y)}{argmax} \sum_{t=1}^{T} \log P(y_t) + \underset{P(y)}{argmax} \sum_{t=1}^{T} \sum_{i=1}^{M} \log P(\phi_i(x_t)|y_t)$$

Was bedeutet z.B. die Linke Seite? $P = \underset{P(y)}{argmax} \sum_{t=1}^{T} \log P(y_t)$
Wenn wir uns sie anschauen, suchen wir hier ein Wahrscheinlichkeitsmodell, was, wenn wir über die Trainingsdaten iterieren, einen höheren summierten Wert liefert als wenn wir ein anderes Modell anschauen.
Beispiel: Man sieht ein Rotes, ein Grünes und ein Blaues Auto. Man nimmt ein Modell an, das man grüne Autos mit 90% Wahrscheinlichkeit sieht, alle anderen mit 5%. Der summierte Wert betrifft daraufhin:
log(0.05)+log(0.05)+log(0.9)=-2.647
Nehmen wir aber 33% für jede Farbe an:
log(0.33)+log(0.33)+log(0.33)=-1.44
Nichts ändert sich, wenn wir stattdessen 2 rote, 2 grüne und 2 blaue Autos

gesehen haben, das 33% Modell wird immer einen höheren Wert erlangen. Die Formel würde dann so aussehen:

2*log(0.33)+2*log(0.33)+2*log(0.33)

Das heisst, wir können die Formel leicht umstrukturieren, indem wir auf 'unterschiedliche? Events und deren Anzahlen achten:

$argmax_p \sum_v count(v) \log P(v)$

Wir wollen, das sich alle Wahrscheinlichkeiten auf 1 aufsummieren, deswegen führen wir einen Lagrangian MUltiplier ein:

$argmax_{p,\lambda} \sum_v count(v) \log P(v) - \lambda(\sum_v P(v) - 1)$

Abgeleitet nach P(v):

$\sum_v \frac{count(v)}{P(v)} - \lambda$

Das auf 0 setzen und umformen:

$\sum_v \frac{count(v)}{P(v)} - \lambda = 0$

$\sum_v \frac{count(v)}{P(v)} = \lambda$

$\sum_v count(v) = \lambda * P(v)$

$\sum_v \frac{count(v)}{\lambda} = P(v)$

Jetzt wissen wir, wie wir die Wahrscheinlichkeit berechnen, wenn wir Wissen was für einen Wert unser lambda annehmen muss.

Wir wissen, dass:

$\sum_v P(v) = 1$

Also können wir nun alles einsetzen:

$\sum_v \frac{count(v)}{\lambda} = 1$

$\sum_v count(v) = lambda$

Lambda ist also alle Counts zusammen genommen.

Also haben wir die endgültige Formel für unsere MLE:

$\sum_v \frac{count(v)}{\sum_{v'} count(v')} = P(v)$

Und damit kommen die Formeln auf Seite 40 der Folien heraus.