

Antwoorden Leren: Schrijfopdracht 3

Semester 1 2014/15

Deadline: 13 November 2014
MOET INDIVIDUEEL GEMAAKT WORDEN

1. (1 punt)

- (a) We passen op een dataset Logistic Regression met en zonder regularisatie toe. Hierbij optimaliseren we de parameter λ . Zijn naar verwachting de parameters θ_i dan bij gebruik van regularisatie dan groter, kleiner of gelijk zijn aan wat we vinden zonder regularisatie?

Regularisatie maakt dat parameterwaarden dicht bij 0 een lagere "cost" hebben dan grotere positieve of lagere negatieve waarden.

- (b) We doen nu hetzelfde, dus Logistic Regression met en zonder regularisatie, maar met een veel grotere dataset uit hetzelfde domein. Zal dit effect hebben op de optimale waarde van λ en de parameters θ ?

Als we meer data hebben dan is regularisatie steeds minder nodig voor het vermijden van overfitting. Als we λ optimaliseren voor grote datasets dan zal de optimale λ steeds kleiner worden.

2. (1 punt) Een bekend punt uit de geschiedenis van de AI is dat een perceptron geen zgn Booleaanse XOR ("exclusive OR") functie kan leren. Ga na of (gewone) Logistische Regressie dat wel kan: experimenteel, door trainingsdata te maken en te kijken wat er gebeurt, analytisch door de "decision boundary" van logistische regressie te bekijken of beide. Inleveren: observaties van experiment en analyse.

Met "gewone" logistische regressie wordt hier bedoeld "zonder extra variabelen, dus zonder bijvoorbeeld kwadraten oid. In dat geval heeft Logistische Regressie een lineaire "class boundary" en kan het dus geen niet-lineaire patronen zoals XOR leren. Als je dit uitprobeert dan zie dat bij data zonder ruis de "cost" niet naar 0 gaat.

3. (1 punt)

We have a neural network with only a single input node X, a hidden layer node Z and an output node Y. There are no bias nodes. The parameter of the link (X, Z) is 0.2 and of link (Z, Y) is 0.1. The input value of a training example is -5. We do not use regularization.

1. What will be the activation of Y?

Input is -5. We krijgen: $a(Z) = 0.27$, $a(Y) = 0.5$

2. Suppose that the activation of Y that was found by forward propagation is equal to 0.8 and the actual value is 1 is. Will the parameters of (X, Z) en (Z, Y) become larger or smaller? Explain your answer in qualitative terms.

Dit betekent dat de output waarde van het netwerk te klein is. Backprop moet dus de parameter van de link (Z, Y) hoger maken. Verder was de $a(Z)$ ook te laag en die moet dus ook verhoogd. NB: $a(Z)$ is altijd tussen 0 en 1.

3. Now calculate the new parameters, using $\alpha = 0.1$, the output value obtained in the first question above, and $Y = 1$ for the input value -5.

$a(Y) = 0.5$. Since $Y = 1$ we get $\delta Y = -0.5$, $\delta(Z) = 0.13$, new $\theta(Z, Y) = 0.11$, and new $\theta(X, Z) = 0.19$