Tentamen Leren (2013-2014)

29 April 2014
18.00 - 20.30 uur

NB: Leg je collegekaart + ID op je tafel en zet je handtekening op het tentamen. Alle vragen wegen tellen evenveel mee in de eindscore.

Vraag 1 (*9 punten*)

Beschouw de volgende dataset:

ID	X1	X2	Y
1	a	g	p
2	a	h	q
3	b	h	p
4	b	g	q
5	c	g	p
6	c	h	p
7	a	g	p
8	b	g	?

1. Wat is de "prior" probability van p, $P(p)$?

2. Is het mogelijk om op basis van de data na te gaan of de Naive Bayes assumptie waar of onwaar voor een dataset? Zoja, hoe? Zonee, waarom niet? .

3. De MAP classificatie van voorbeeld 8 op basis van Naive Bayes is "p". Laat dit zien.

4. De classificatie van voorbeeld 8 volgens 1-nearest neighbour is "q" (met voorbeeld 4). Wat doen we hiermee? Wie zou er gelijk hebben?

Vraag 2 (*9 punten*)

1. Geef de vorm van de logistic regression classifier als functie van parameters θ dus geef de functie $f(\theta, x)$ en de drempel voor een beslisregel in de vorm $ALS\ f(\theta,x)\ >\ drempel\ DAN\ y = 1\ ANDERS\ y = 0$.

2. We kunnen ook een beslisregel formuleren door uit te gaan van $\frac{P(y=1|x)}{P((y=0|x)}$: ALS $\frac{P(y=1|x)}{P((y=0|x)} > 1$ DAN $y = 1$ en ANDERS $y = 0$. Vergelijk deze regel met de beslisregel *ALS* $\theta_0 + \theta^T.x < 0$ DAN $y = 1$ en anders $y = 0$. Tip: $P(y = 1|x) = 1 - P(y = 0|x)$. Neem aan beide kanten de natuurlijke logarithme van de beslisregel uit de vorige deelvraag.

Vraag 3 (*9 punten*)

1. Wat wordt bedoeld met de term ”gradient” in de context van ”gradient descent”? Geef een definitie en een formule voor regressie op basis van de ”mean squared error”.

2. Stel dat we multipele regressie willen doen met variabelen X1 .. Xn en behalve de variabelen willen we ook de kwadraten van die variabelen gebruiken. We krijgen dan 2n voorspellende variabelen. Leid een gradient en een update functie af voor een gradient descent algorithme voor deze regressiefunctie.

Vraag 4 (*10 punten*)
We hebben een neuraal netwerk met alleen een inputknoop X, een hidden layer knoop Z en een outputknoop Y. Het gewicht van link (X, Z) is 0.2 en van link (Z, Y) is -0.1. De input van een trainingsvoorbeeld is -10. Er zijn geen ”bias” knopen en we doen geen regularisatie.

1. Wat wordt de activatie van Y? (Reken dit uit of geef anders een formule waarin zo min mogelijk operaties zitten.)

2. Als Y van dit trainingsvoorbeeld 1 is, zullen de gewichten van (X, Z) en (Z, Y) dan groter of kleiner worden. Licht je antwoord toe.

3. Wat is in dat geval de (teruggepropageerde) error van Z?

Vraag 5 (*9 punten*)
Stel dat we voor een dataset multipele logistische regressie en dat we waarden van de parameters θ_0, θ_1 en θ_2 vinden waarvoor $J(\theta_0, \theta_1, \theta_2) = 0$. Welke van de onderstaande uitspraken moet(en) dan waar zijn? Licht je antwoord kort toe.

1. Dit is niet mogelijk. Door de definitie van $J(\theta_0, \theta_1, \theta_2)$, is het niet mogelijk dat er een $\theta_0, \theta_1, \theta_2$ zijn zo dat $J(\theta_0, \theta_1, \theta_2) = 0$

2.) Hiervoor moet $y(i) = 0$ voor elke waarde $i = 1, 2, , m$.

3. Waarschijnlijk is gradient descent op een lokaal minimum gekomen en is niet het globale minimum gevonden.

4. Het model met de gevonden waarden overfit de data.

5. Als voor deze waarden van $\theta_0, \theta_1, \theta_2$ de waarde van $J(\theta_0, \theta_1, \theta_2)$, gelijk is aan 0, dan moet $h_\theta(x(i)) = y(i)$ voor elk trainingsvoorbeeld $(x(i), y(i))$

Vraag 6 (*10 punten*)
Als we een leeralgorithme toepassen op data en dit maakt een classifier, dan kunnen fouten die de classifier maakt op nieuwe data het gevolg zijn van "variance" of van "bias", of beide. We kunnen een "regularizer" toevoegen aan de doelfunctie (resp. de "loss") met een bepaald gewicht en een variant van het algorithme afleiden.

1. Geef een voorbeeld van een loss function en een regularizer voor multipele lineaire regressie, neurale netwerken.

2. We kunnen regularisatie ook toepassen op beslisbomen. Hoe?

3. Heeft het toevoegen van een regularizer effect op de variance component in de error? Op de bias? Op beide?