

STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Beräkningsmatematik

TENTAMEN
Introduktion till maskininläring
19 april 2021

Tentamen för kursen
Introduktion till maskininläring
19 april 2021 8–14

Ansvarig lärare: Sebastian Rosengren, rosengren@math.su.se

Examinator: Lars Arvestad

Hjälpmedel: Kurslitteratur och föreläsningssanteckningar.

Återlämning: Information om återlämning skickas ut via kursforum.

Lösningar: Finns på kursens hemsida efter skrivtidens slut.

Varje korrekt löst uppgift ger 12 poäng. E: 30p, D: 35p, C: 40p, B: 45p, A: 50p.

Svar ska motiveras och resonemang skall vara klara och tydliga att följa.

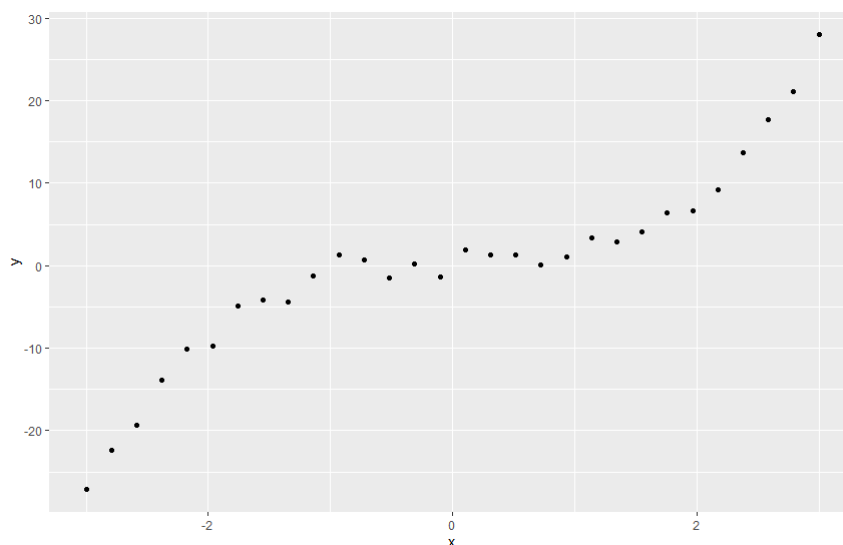
Vidare är det en god idé att läsa bifogad fil:

Praktiskt information angående inlämning av tentamen.

Lycka Till!

Uppgift 1

Antag att vi har följande 30 datapunkter.



Antag vidare att $y_i = f_\theta(x_i) + \epsilon_i$, där ϵ_i , $i = 1, 2, \dots, 30$ är oberoende slumpfel. Ange för nedanstående modeller om de har hög eller låg *bias*, samt hög eller låg *varians*. Endast svar räcker.

a.) $f_\theta(x) = \alpha + \sum_{i=1}^{25} \beta_i x^i$ (2 p)

b.) $f_\theta(x) = \alpha + \sum_{i=1}^3 \beta_i x^i$ (2 p)

c.) $f_\theta(x)$ är ett neuralt nät med ReLU aktivering, sju gömda lager, med fem noder per lager. (2 p)

d.) $f_\theta(x) = \sum_{m=1}^{30} c_m 1\{x \in R_m\}$, alltså ett beslutsträd med 30 regioner. (2 p)

e.) $f_\theta(x) = \frac{1}{2} \sum_{x_i \in N_2(x)} y_i$, alltså en knn regression med $k = 2$. (2 p)

f.) I termer av bias varians vad är skillnaden mellan modellen i d.) och modellen i d.) tränad med bagging. (2 p)

Uppgift 2

a.) Definera cross-entropy för modellen enkel logistisk regression. (2 p)

b.) Vad är det *primära* syftet med boosting. Vilka egenskaper bör en modellfamilj ha för att boosting ska vara lämpligt. (2 p)

- c.) Vad är det primära syftet med PCA inom klustering. (2 p)
- d.) Definiera vad som menas med en Markov policy π . Varför är just den typen av policies viktiga inom reinforcement learning? (2 p)
- e.) Beskriv två situationer då value iteration inte kan användas för att hitta en optimal policy. (2 p)
- f.) Låt $f(x) = w^T \phi(x) + b$ vara en support vector machine för binär klassifikation. Hur väljs parametern w^t ? (2 p)

Uppgift 3

Läs in data från filen `data_uppgift3.csv`. Data består av output y och input $x = (x_1, x_2, x_3)$.

- a.) Anpassa (den linjära regressions) modellen $f_1(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ på hela datasetet, och ange parameterskattningarna. (4 p)
- b.) Dela nu upp data i en 75/25 tränings/test-split. Anpassa ovanstående modell på träningsdata och beräkna *mean absolute error* för modellen på testdata. (4 p)

c.) Låt

$$f_2(x) = \alpha + \beta_1 x_1$$

$$f_3(x) = \alpha + \beta_2 x_2$$

$$f_4(x) = \alpha + \beta_3 x_3.$$

I termer av *mean absolute error* vilken av modellerna $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_4(x)$ verkar generalisera bäst till ny data? Motivera. (4 p)

Uppgift 4

Låt $f_\theta(x)$ vara ett neuralt nätverk med ReLu aktivering för binär klassifikation. Antag vidare att $x \in \mathbb{R}^{10}$ (tiodimensionell input), samt att $f_\theta(x)$ har 5 gömda lager, med 10, 12, 14, 16, respektive 18 noder.

- a.) Antag att vi har tillräckligt med träningsdata för att ovanstående modell kan vara lämplig. Ange för- och nackdelar med ovanstående modell jämfört med en logistisk regressionsmodell. (4 p)
- b.) Beräkna hur många parametrar det finns i modellen. Ange hur outputen $f_\theta(x)$ beräknas givet att vi har värdet på det sista gömda lagret $h^{(5)}$. (4 p)
- c.) Vad används early stopping till? Skissa algoritmen. (4 p)

Uppgift 5

a.) Skriv ut formeln $p(x)$ för en blandad normalfördelning med två komponenter och parametrar $\mu_1, \mu_2, \Sigma_1, \Sigma_2, \pi_1$. Beskriv vad parametrarna har för betydelse. (2 p)

b.) För datasetet nedan, kommer klustring med k -means och en Gaussian mixture model (GMM) med EM-algorithmen resultera i samma klusterindelning? Motivera. (4 p)



c.) I första iterationen av GMM med EM-algoritmen väljer vi μ_1 och μ_2 enligt bilden nedan, samt $\pi_1 = 0.5$. I vilken riktning kommer μ_1 och μ_2 röra på sig? (2 p)



d.) I samma iteration av EM-algoritmen som i c), kommer likelihooden minska eller öka? Förklara. (2 p)

e.) Om vi itererar EM-algoritmen i c) flera gånger, kommer värdet på π_1 minska eller öka (från sitt initiala värde)? Förklara. (2 p)