

Räkneövning 2

1. Säg att diskriminantfunktionerna för ω_1 och ω_2 är $g_1(\mathbf{x}) = \mathbf{w}_1^T \mathbf{x} + w_{1,0}$ respektive $g_2(\mathbf{x}) = \mathbf{w}_2^T \mathbf{x} + w_{2,0}$.
 - (a) Vad är ekvationen för beslutsytan? Skriv den på formen $g(\mathbf{x}) = 0$.
 - (b) Rita upp beslutsytan för fallet $\mathbf{w}_1 = (1, 1)^T$, $w_{1,0} = 2$, samt $\mathbf{w}_2 = (0, 2)^T$, $w_{2,0} = -1$. Markera dessutom på vilka sidor om gränsen som besluten " $\mathbf{x}$ tillhör ω_1 ", respektive " $\mathbf{x}$ tillhör ω_2 " fattas.
2. Visa att viktvektorer som är optimala i avseendet att de minimerar summan av kvadratfel, asymptotiskt går mot "Mean squared error"-lösningen.
3. Vid logistisk klassning modellerar vi logaritmen av kvoten $P(\omega_i|\mathbf{x})/P(\omega_M|\mathbf{x})$ som

$$\ln \frac{P(\omega_i|\mathbf{x})}{P(\omega_M|\mathbf{x})} = w_{i,0} + \mathbf{w}^T \mathbf{x}, \quad \text{för } i = 1, \dots, M-1 \quad (1)$$

Via bivillkoret $\sum_{i=1}^M P(\omega_i|\mathbf{x}) = 1$ får vi att

$$P(\omega_M|\mathbf{x}) = \frac{1}{1 + \sum_{i=1}^{M-1} \exp(w_{i,0} + \mathbf{w}^T \mathbf{x})}. \quad (2)$$

Visa att logaritmen av kvoten $P(\omega_i|\mathbf{x})/P(\omega_M|\mathbf{x})$ får precis denna struktur om vi istället utgår från antagandet att klasserna är normalfördelade med olika väntevärdesvektorer, $\mathbf{m}_1, \dots, \mathbf{m}_M$, fast med identiska kovariansmatriser.

4. Anta att du skapar en *support vector machine* (SVM) med hjälp av följande två dataset: $X_1 = \{(0, 0)^T, (1, 0)^T, (0, 1)^T, (1, 1)^T\}$ och $X_2 = \{(2, 2)^T, (3, 2)^T, (2, 3)^T, (3, 3)^T\}$.
 - (a) Var kommer beslutsgränsen för denna SVM gå? Motivera varför. (Vilket optimeringsproblem försöker vi lösa?)
 - (b) Vilka av vektorerna ovan är s.k. supportvektorer?
5. Livet i den lilla byn var tråkigt och det enda nöjet var att räkna bilar och spionera på grannarna. Byns ingenjör ville förenkla detta jobb och satte upp en mätstation utanför sitt hus. Denna mätte dels hastigheten och uppskattade dels vikten hos den förbipasserande bilen. Vid kalibreringen noterade han vilken av hans tre grannar som körde och han förde in dessa data i en tabell samt ett s.k. spridningsdiagram, se nedan.
 - (a) En dag registrerade mätstationen en bil med hastigheten 80 km/h och vikten 1000 kg . Vilken av grannarna var det? Fatta beslutet m.h.a KNN med $K = 3$ och $K = 5$ samt $K = 15$. Kommentarer?
 - (b) Du noterar kanske att det inte är helt uppenbart hur man ska mäta avstånd mellan mönstervektorerna. Ge ett förslag på hur man kan se till att inverkan från hastigheten och vikten blir ungefärligt likvärdiga.

Karl	Johan	Nils
1030/100	950/90	870/50
1060/105	1000/70	930/70
1020/85	1010/85	950/80
1010/75	940/50	900/85
990/95	970/80	
930/95		

Table 1:

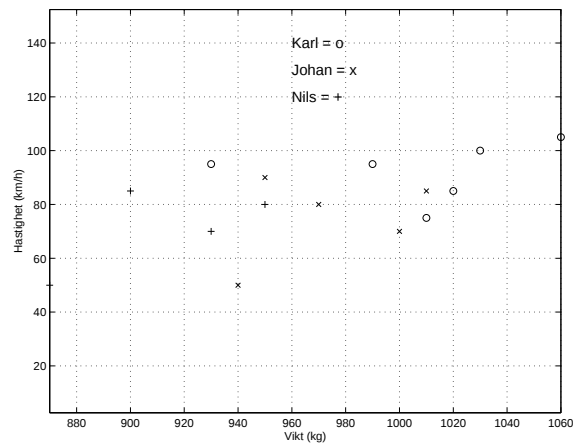


Figure 1:

6. Backpropagation. Betrakta ett 3-lagernät med strukturen 2-2-1 (två neuroner i lager 1 osv) och där indatavektorn, $\mathbf{x}$, har dimensionen 2. Låt $g(\mathbf{x}_n, \theta)$ beteckna utsignalen från nätet (skalär eftersom vi har en neuron i sista lagret). Vid träningen ska vi minimera kriteriet $J = \sum_n (t_n - g(\mathbf{x}_n, \theta))^2$ där vektorn θ representerar alla parametrar som ingår i nätet.
 - (a) Rita upp MLPn och markera i figuren vilka parametrar (vikter) som ingår.
 - (b) Skriv upp $g(\mathbf{x}, \theta)$ som en funktion av dessa parametrar. ($\mathbf{x} = (x_1, x_2)^T$)
 - (c) Beräkna de partiella derivatorna m.a.p. samtliga dessa parametrar.
 - (d) Förklara utifrån dessa derivator namnet *error backpropagation*.
7. Träning av MLP sett som ML-skattning. Anta att vi vid träning av en MLP (som vanligt) minimerar summan av kvadratfel för träningsdata. Visa att detta motsvarar ML-skattning av parametrarna (vikter, bias) om vi ställer upp en modell för sambandet mellan x och y :

$$y = f(x, \mathbf{w}) + e, \quad (3)$$

där $f(x, \mathbf{w})$ är utsignalen från nätet som beror av parametervektorn $\mathbf{w}$ (alla vikter i nätet) och e är en okänd mätstörning som antas vara normalfördelad $e \sim N(0, \sigma^2)$.

8. Olinjäriteten $\phi(\cdot)$ som ingår i modellen av en neuron brukar ofta väljas som den logistiska funktionen: $\phi(a) = \frac{1}{1+\exp(-a)}$. Visa att detta val leder till att derivatan $\phi'(a)$ kan beräknas direkt utifrån värdet på $\phi(a)$.
9. Om man har ett klassningsproblem så kan man, som ett alternativ till den kvadratiske funktionen i uppg. 6, använda sig av *korsentropi* som felkriterium. Låt $y_k(i)$ vara mönster nr i 's indikatorvärde för klass nr k . M.a.o $y_k(i) = 1$ om $\mathbf{x}_i \in \omega_k$ och noll annars. För M klasser och N exempel definieras korsentropin som

$$J = - \sum_{i=1}^N \sum_{k=1}^M (y_k(i) \ln \hat{y}_k(i) + (1 - y_k(i)) \ln(1 - \hat{y}_k(i))), \quad (4)$$

där $\hat{y}_k(i)$ är utsignalen från tex MLPn. Plotta en term i denna kriteriefunktion som funktion av $\hat{y}_k(i)$ och jämför med motsvarande plot för en kvadratisk felfunktion. Obs, du har två fall: (a) $y_k(i) = 1$ och (b) $y_k(i) = 0$. Slutsatser?

10. Ytterligare en förekommande variant på det vanliga kvadratiske felkriteriet är $J = \sum_{n=1}^N |t_n - g(\mathbf{x}_n, \theta)|^\beta$ där θ är den eftersökta parametervektorn, $|\cdot|$ är absolutbeloppet och parametern $\beta > 0$. (För $\beta = 2$ får vi det vanliga kvadratfelet.) Vi har tidigare sett att man kan tolka minimering av summan av kvadratfelen som en ML-skattning om vi antar en modell

$$t_n = g(\mathbf{x}_n, \theta) + e_n, \quad (5)$$

där $e_n \sim N(0, \sigma^2)$, dvs e_n är normalfördelad.

- (a) Utför en liknande tolkningsövning för $\beta \neq 2$. Alltså, vilket antagande om fördelningen för e_n leder i samband med ML-skattning till ovanstående kriterium J ?
- (b) Rita upp $p(e_n)$ för fallet $\beta = 1$.
- (c) Rita upp $p(e_n)$ för fallet $\beta = \infty$.
11. Vi använder ett beslutsträd för att lösa ett tvåklassproblem. En av noderna innehåller 6 st. ex från ω_1 och 4 från ω_2 .
 - (a) Vilken är nod-orenheten (node impurity, se ekv. (4.70) T-K)?
 - (b) Två alternativa uppdelningar av noden föreslås: (A) t_Y har 2 ex från ω_1 och 1 ex från ω_2 . (B) t_Y har 3 ex från ω_1 och 0 ex från ω_2 . Vilken av (A) och (B) bör vi välja?