1 Uppgifter Räkneövning 5

1. Anta att du har tillgång till ett stort antal sekvenser som har genererats från samma HMM. Förklara i detalj hur du kan utnyttja Viterbialgoritmen för att skatta parametrarna i modellen. (Viterbi-learning).

2. I ekv. (6.37) i Rabiner påstås (uttryckt med våra beteckningar) att

$$\frac{P(s^t = q_i, s^{t+1} = q_j, X^1, ..., X^T|\omega)}{P(\mathcal{O}|\omega)} = \frac{\alpha_i(t) t_{ji} e_{jX^{t+1}} \beta_j(t+1)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(t) t_{ji} e_{jX^{t+1}} \beta_j(t+1)} \tag{1}$$

   där, som vanlig, $\alpha_i(t) = P(X^1, ..., X^t, s^t = q_i|\omega)$ och $\beta_j(t) = P(X^1, ..., X^{t+1}|s^t = q_i, \omega)$ och N är antalet tillstånd i modellen.

   (a) Visa att påståendet är sant.

   (b) Förklara i detalj hur ovanstående samband kommer in vid ML-skattning av t_{ji}.

   (c) Förklara i detalj hur ovanstående samband kan användas vid ML-skattning av e_{iX^t}. Alternativt, förklara hur du kan utnyttja produkten $\alpha_i(t)\beta_i(t)$ vid samma skattning.

3. Civert har tänkt lösa ett funktionsapproximationsproblem mha en MLP. Funktionen som han vill approximera mha träningsdata är 1-D till 1-D och han har beslutat att välja strukturen (1)-M_1-M_2-M_3-1, dvs M_1 neuroner i första lagret, M_2 i andra, M_3 i tredje och 1 neuron i sista lagret. Han väljer funktionerna $\phi(\cdot)$ att vara: logistisk i första lagret, linjär i andra och tredje lagret. Övertyga Civert om att han ett gjort fel.

4. Man behöver inte nödvändigtvis använda ett kvadratisk felkriterium vid träning av MLP eller RBF etc. En förekommande variant är kriteriet $J = \sum_{n=1}^N |t_n - g(\mathbf{x}_n, \theta)|^\beta$ 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

$$t_n = g(\mathbf{x}_n, \theta) + e_n, \tag{2}$$

   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 $\beta = \infty$.