These notes are intended to give a summary of relevant concepts from the lectures which are helpful to complete the tutorial sheet. It is not intended to cover the lectures thoroughly. Learning this content is not a replacement for working through the lecture material and the tutorial sheet.

**Markov chains** — A distribution factorised such that each variable \( x_i \) depends on \( L \) previous (contiguous) nodes \( \{x_{i-L}, \ldots, x_{i-1}\} \)

\[
p(x_1, \ldots, x_d) = \prod_{i=1}^d p(x_i \mid x_{i-L}, \ldots, x_{i-1})
\]

For \( L = 1 \) we have a 1\(^{st}\)-order Markov chain, \( p(x_1, \ldots, x_d) = \prod_{i=1}^d p(x_i \mid x_{i-1}) \)

The transition distribution \( p(x_i \mid x_{i-1}) \) gives the probability of transitioning to different states. However, if this does not depend on \( i \), then the Markov chain is said to be homogeneous.

**Hidden Markov model (HMM)** — A 1\(^{st}\)-order Markov chain on latent variables \( h_i \) (hiddens), with an additional set of visible varaibles \( v_i \) that represent observations. An emission distribution \( p(v_i \mid h_i) \) gives the probabilities of the observations \( v_i \) (visibles) taking different values, if the observations are real-valued then \( p(v_i \mid h_i) \) will be a probability density function.

\[
p(h_{1:d}, v_{1:d}) = p(v_1 \mid h_1)p(h_1)\prod_{i=2}^d p(v_i \mid h_i)p(h_i \mid h_{i-1})
\]

An HMM is said to be stationary if its transition and emission distributions don’t depend on \( i \).

**Alpha-recursion** A recursive process that propagates information forwards, from \( h_{s-1} \) to \( h_s \)

\[
\alpha(h_s) = p(v_s \mid h_s) \sum_{h_{s-1}} p(h_s \mid h_{s-1})\alpha(h_{s-1}) \tag{1}
\]

\[
\alpha(h_1) = p(h_1)p(v_1 \mid h_1) \propto p(h_1 \mid v_1) \tag{2}
\]

**Beta-recursion** A recursive process that propagates information backwards, from \( h_{s+1} \) to \( h_s \)

\[
\beta(h_s) = \sum_{h_{s+1}} p(h_{s+1} \mid h_s)p(v_{s+1} \mid h_{s+1})\beta(h_{s+1}) \tag{3}
\]

\[
\beta(h_u) = 1 \tag{4}
\]
**Filtering** — Given previous observations $v_{1:t-1}$, and the current observation $v_t$, infer the current hidden state at time $t$

$$p(h_t \mid v_{1:t})$$ (5)

**Smoothing** — Given previous observations $v_{1:t-1}$, and some future observations $v_{t+u}$, infer the hidden state at time $t$

$$p(h_t \mid v_{1:u})$$ (6)

**Prediction** — Given some previous observations $v_{1:u}$, infer the hidden state at time $t$

$$p(h_t \mid v_{1:u})$$ (7)

**Most likely hidden path (Viterbi alignment)** — Given previous observations $v_{1:t-1}$, and the current observation $v_t$, find the most likely hidden path

$$\arg\max_{h_{1:t}} p(h_{1:t} \mid v_{1:t})$$ (8)