Patrick Pegus
Mini Project 2
November 7, 2015
CMPSCI-689
Prof. Sridhar Mahadevan

1. (a) The state variables at time t are marginally independent because the observation at that time d-separates them. In other words, $P(S(t,1)|S(t,2)) = P(S(t,1))$ because $Y(t)$ is a collider node in their path. However, $P(S(t,1)|S(t,2)) \neq P(S(t,1))$ when $Y(t)$ then d-connects them. The state variables at time t are conditionally independent of the past history of state variables given the state variables at $t-1$ because those given variables d-separate them from past states.

   (b) To convert the factorial HMM to a regular HMM, collapse states $S(t,1), \ldots S(t,M)$ to a single state $S(t)$ that has K^M values, which is enough to represent all possible state combinations of the former states. Since the time complexity of the forward algorithm on an HMM is $O(L^2T)$ where L is the number of state values, the complexity of the converted HMM is $O\left(\left(K^M\right)^2 T\right) = O(K^{2M}T)$.

2. (a) **Lagrange dual**

$$L(w,\xi,\alpha) = \lambda\|w\|^2 + \sum_{i=1}^{l} \xi_i^2 + \sum_{i=1}^{l} \alpha_i(y_i - \langle w, x_i\rangle - \xi_i)$$

$$\max_{\alpha} \left(L_D(\alpha) = \min_{w,\xi} L(w,\xi,\alpha) \right)$$

Assure 0 duality gap

$$0 = \frac{\delta}{\delta w} L(w,\xi,\alpha) = 2\lambda w - \sum_{i=1}^{l} \alpha_i x_i \iff w = \frac{1}{2\lambda}\sum_{i=1}^{l} \alpha_i x_i$$

$$0 = \frac{\delta}{\delta \xi_k} L(w,\xi,\alpha) = 2\xi_k - \alpha_k \iff \xi_k = \frac{\alpha_k}{2}$$

$$\begin{aligned}
L_D(\alpha) &= \lambda\|\frac{1}{2\lambda}\sum_{i=1}^{l} \alpha_i x_i\|^2 + \sum_{i=1}^{l}\left(\frac{\alpha_i}{2}\right)^2 + \sum_{i=1}^{l} \alpha_i(y_i - \langle \frac{1}{2\lambda}\sum_{j=1}^{l} \alpha_j x_j, x_i\rangle - \frac{\alpha_i}{2}) \\
&= \frac{1}{4\lambda}\|\sum_{i=1}^{l} \alpha_i x_i\|^2 + \sum_{i=1}^{l} \frac{\alpha_i^2}{4} + \sum_{i=1}^{l} \alpha_i y_i - \frac{1}{2\lambda}\sum_{i=1}^{l} \alpha_i \langle \sum_{j=1}^{l} \alpha_j x_j, x_i\rangle - \sum_{i=1}^{l} \frac{\alpha_i^2}{2} \\
&= \frac{1}{4\lambda}\sum_{i=1}^{l} \alpha_i \langle \sum_{j=1}^{l} \alpha_j x_j, x_i\rangle - \sum_{i=1}^{l} \frac{\alpha_i^2}{4} + \sum_{i=1}^{l} \alpha_i y_i - \frac{1}{2\lambda}\sum_{i=1}^{l} \alpha_i \langle \sum_{j=1}^{l} \alpha_j x_j, x_i\rangle \\
&= \sum_{i=1}^{l} \alpha_i y_i - \frac{1}{4\lambda}\sum_{i=1}^{l} \alpha_i \langle \sum_{j=1}^{l} \alpha_j x_j, x_i\rangle - \sum_{i=1}^{l} \frac{\alpha_i^2}{4} \\
&= \sum_{i=1}^{l} \alpha_i y_i - \frac{1}{4\lambda}\sum_{i=1}^{l} \alpha_i \sum_{j=1}^{l} \alpha_j \langle x_j, x_i\rangle - \sum_{i=1}^{l} \frac{\alpha_i^2}{4}
\end{aligned}$$

(b) Solution to kernel ridge regression occurs when $\frac{\delta}{\delta\alpha} L_D(\alpha) = 0$.

$$
\begin{aligned}
0 &= \frac{\delta}{\delta\alpha} L_D(\alpha) = \ldots \text{working backwards, but can't figure out this step} \ldots = y - \frac{G\alpha}{2\lambda} - \frac{\alpha}{2} \\
0 &= y - (G + \lambda I)\,\frac{\alpha}{2\lambda} \\
(G + \lambda I)\,\frac{\alpha}{2\lambda} &= y \\
\frac{\alpha}{2\lambda} &= (G + \lambda I)^{-1}\, y \\
\alpha &= 2\lambda\,(G + \lambda I)^{-1}\, y
\end{aligned}
$$