Ex 6

1.1) Show that $F(x,y) = F_1(x,y) + F_2(x,y)$

$$F(x,y) = \sum_i \sum_{\bar{y}} \alpha_i(\bar{y}) \langle \Phi(x_i,\bar{y}), \Phi(x,y) \rangle \quad \text{—} \; 1$$

--- Eqn (8)

$$\langle \Phi(x_i,\bar{y}), \Phi(x,y) \rangle = \sum_{s,t} [[\bar{y}^{s-1} = y^{t-1} \wedge \bar{y}^{s} = y^{t}]]$$

$$+ \sum_{s,t} [[\bar{y}^{s} = y^{t}]] \, k(x_i^s, x^t) \quad \text{—} \; 2$$

--- using Eqn (7)

(note that the variables are consistent with Eqn 8 but slightly different from Eqn 7)

$$F_1(x,y)$$

$$= \sum_{\sigma,\tau} \underbrace{\boxed{\left(\sum_{i,\bar{y}} \alpha_i(\bar{y}) \sum_t [[\bar{y}^{t-1} = \sigma \wedge \bar{y}^{t} = \tau]] \right)}}_{\text{Transition matrix}} \cdot \sum_s [[y^{s-1} = \sigma \wedge y^{s} = \tau]]$$

--- Eqn (9a, 9b)

Viterbi decoding

Emission matrix

$$F_2(x,y)$$

$$= \sum_{\tau,\sigma} [[y^s = \sigma]] \boxed{\sum_{i,t} \left(\sum_y [[y^t = \sigma]] \, \alpha_i(y) \right) \cdot k(x^s, x_i^t)}$$

--- Eqn (10a, 10b).

with $\sigma, \tau, y^t \in \Sigma$: "set of all possible labels"

Now, using ① & ②

$$F(x,y) = \sum_{i,\bar{y}} \alpha_i(\bar{y}) \sum_{s,t} \underbrace{[[\bar{y}^{s-1} = y^{t-1} \wedge \bar{y}^s = y^t]]}_{A}$$

$$+ \sum_{i,\bar{y}} \alpha_i(\bar{y}) \sum_{s,t} \underbrace{[[\bar{y}^s = y^t]]}_{B} k(x_i^s, x^t)$$

$$= \underbrace{\sum_{i,\bar{y},s,t} \alpha_i(\bar{y}) \cdot A}_{F_1'} + \underbrace{\sum_{i,\bar{y},s,t} \alpha_i(\bar{y}) \cdot B \cdot k(x_i^s, x^t)}_{F_2'}$$

(since $\alpha_i(\bar{y})$ is constant with respect to 's', 't')

So, it is enough to show that $F_1' = F_1$ & $F_2' = F_2$.

$$B = [[\bar{y}^s = y^t]] = \sum_{\sigma} [[\bar{y}^s = \sigma]] [[\bar{y}^s = y^t]]$$

this makes sense because $\bar{y}^s$ is equal to one particular value from the set of possible labels, therefore $[[\bar{y}^s = \sigma]] = 1$ for only one (the correct) label in this summation and is 0 for all the remaining terms (corresponding to the other labels)

Thus,

$$B = [[\bar{y}^s = y^t]] = \sum_\sigma [[\bar{y}^s = \sigma]][[\bar{y}^s = y^t]]$$

$$= \sum_\sigma [[\bar{y}^s = \sigma \wedge \bar{y}^s = y^t]] = \sum_\sigma [[\bar{y}^s = \sigma \wedge y^t = \sigma]]$$

$$= \sum_\sigma [[\bar{y}^s = \sigma]]\,[[y^t = \sigma]]$$

Then,

$$F_2' = \sum_{i,\bar{y},s,t} \alpha_i(\bar{y}) \left[\sum_\sigma [[\bar{y}^s = \sigma]]\,[[y^t = \sigma]] \right] k(x_i^s, x^t)$$

$$= \sum_{i,\bar{y},s,t,\sigma} \alpha_i(\bar{y})\,[[\bar{y}^s = \sigma]]\,[[y^t = \sigma]]\,k(x_i^s, x^t)$$

$$= \sum_{t,\sigma} [[y^t = \sigma]] \sum_{i,\bar{y},s} \left\{ [[\bar{y}^s = \sigma]]\,\alpha_i(\bar{y}) \cdot k(x_i^s, x^t) \right\}$$

$$= \sum_{t,\sigma} [[y^t = \sigma]] \sum_{i,s} \left(\sum_{\bar{y}} [[\bar{y}^s = \sigma]]\,\alpha_i(\bar{y}) \right) k(x_i^s, x^t)$$

$$= F_2 \quad //$$

(exchanging indices 's' and 't' and using variable 'y' instead of $\bar{y}$ in the summation makes no difference, as discussed in class).

Now,

$$A = [[\bar{y}^{s-1} = y^{t-1}]]\,[[\bar{y}^{s} = y^{t}]]$$

$$= \left(\sum_{\sigma} [[\bar{y}^{s-1} = \sigma]]\,[[y^{t-1} = \sigma]]\right)\left(\sum_{\tau} [[\bar{y}^{s} = \tau]]\,[[y^{t} = \tau]]\right)$$

→ same logic as used for simplifying 'B'.

$$= \sum_{\sigma,\tau} [[\bar{y}^{s-1} = \sigma \wedge \bar{y}^{s} = \tau]]\,[[y^{t-1} = \sigma \wedge y^{t} = \tau]]$$

Thus,

$$F_1' = \sum_{i,\bar{y},s,t} \alpha_i(\bar{y}) \left(\sum_{\sigma,\tau} [[\bar{y}^{s-1} = \sigma \wedge \bar{y}^{s} = \tau]]\,[[y^{t-1} = \sigma \wedge y^{t} = \tau]]\right)$$

$$= \sum_{\sigma,\tau} \left(\sum_{i,\bar{y},s} \alpha_i(\bar{y})\,[[\bar{y}^{s-1} = \sigma \wedge \bar{y}^{s} = \tau]]\right) \sum_{t} [[y^{t-1} = \sigma \wedge y^{t} = \tau]]$$

(exchanging indices 's' & 't')

$$= F_1$$

//

$$\therefore F = F_1' + F_2' = F_1 + F_2$$

1.2) By Eqn (6), $\underline{\Phi}(x,y) = \sum_{t=1}^{T} \underline{\Phi}(x,y;t)$

Also, "all the features extracted at location t are simply stacked together to form $\underline{\Phi}(x,y;t)$.

Thus,

$$\underline{\Phi}(x,y;t) = \begin{pmatrix} \vdots \\ \phi^{tt}_{r\sigma} = [[y^t = \sigma]]\, \psi_r(x^t) \\ \vdots \\ \phi^{(t-1)t}_{\sigma\tau} = [[y^{t-1} = \sigma \wedge y^t = \tau]] \\ \vdots \end{pmatrix} \begin{matrix} \} \; d \times |\Sigma| \\ \\ \} \; |\Sigma| \times |\Sigma| \end{matrix}$$

Then, $\langle \underline{\Phi}(x,y), \underline{\Phi}(\bar{x},\bar{y}) \rangle$

$$= \left\langle \sum_s \underline{\Phi}(x,y;s), \sum_t \underline{\Phi}(\bar{x},\bar{y};t) \right\rangle$$

$$= \sum_s \left\langle \underline{\Phi}(x,y;s), \sum_t \underline{\Phi}(\bar{x},\bar{y};t) \right\rangle$$

$$= \sum_{s,t} \left\langle \underline{\Phi}(x,y;s), \underline{\Phi}(\bar{x},\bar{y};t) \right\rangle$$

Now, let's look at the term inside the summation.

Suppose, $\left\langle \underline{\Phi}(x, y; s), \underline{\Phi}(\bar{x}, \bar{y}; t) \right\rangle = P_1 + P_2$.

such that,

$$P_2 = \sum_{\sigma} \sum_{\gamma} [[y^s = \sigma]] \, \psi_\gamma(x^s) \cdot [[\bar{y}^t = \sigma]] \, \psi_\gamma(\bar{x}^t)$$

$$= \underbrace{[[y^s = \bar{y}^t]]}_{} \sum_{\gamma} \psi_\gamma(x^s) \cdot \psi_\gamma(\bar{x}^t)$$

(same logic as used for A and B in 1.1>)

$$= [[y^s = \bar{y}^t]] \left\langle \psi(x^s), \psi(\bar{x}^t) \right\rangle$$

$$= [[y^s = \bar{y}^t]] \cdot k(x^s, \bar{x}^t) \quad , \quad \text{and}$$

$$P_1 = \sum_{\sigma} \sum_{\tau} [[y^{s-1} = \sigma \wedge y^s = \tau]] [[y^{t-1} = \sigma \wedge y^t = \tau]]$$

$$= [[y^{s-1} = y^{t-1}]] [[y^s = y^t]]$$

$$\Rightarrow \left\langle \underline{\Phi}(x, y), \underline{\Phi}(\bar{x}, \bar{y}) \right\rangle = \sum_{s,t} (P_1 + P_2)$$

$$= \sum_{s,t} P_1 + \sum_{s,t} P_2$$

2.2) Primal (with l_2 penalties) ... Eqn 18

$$\min \frac{1}{2}\|\underset{\sim}{w}\|^2 + \frac{C}{2}\sum_i \xi_i^2$$

such that $z_i(\underset{\sim}{y})\left(\langle \underset{\sim}{w}, \Phi(\underset{\sim}{x}_i, \underset{\sim}{y})\rangle + \theta_i\right) \geq 1 - \xi_i$

$\xi_i \geq 0 \quad \forall\, i = 1(1)n, \ \forall\, \underset{\sim}{y} \in \mathcal{Y}$

<u>Lagrangian</u>:

$$L = \frac{1}{2}\|\underset{\sim}{w}\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2$$

$$- \sum_{i=1}^{n}\sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\left(z_i(\underset{\sim}{y})\left(\underset{\sim}{w}^T\Phi(\underset{\sim}{x}_i, \underset{\sim}{y}) + \theta_i\right) - 1 + \xi_i\right)$$

such that $\alpha_i(\underset{\sim}{y}) \geq 0 \quad \forall\, i = 1(1)n \quad \forall\, \underset{\sim}{y} \in \mathcal{Y}$

$\rightarrow$ maximise w.r.t. α and minimise w.r.t. $\underset{\sim}{w}$, ξ

$$\frac{\partial L}{\partial \xi_i} = C\cdot\xi_i - \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) = 0$$

$$\Rightarrow \xi_i = \frac{1}{C}\sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) \geq 0 \qquad (1)$$

so we don't need to enforce positivity of ξ_i

$$\frac{\partial L}{\partial \underset{\sim}{w}} = \underset{\sim}{w} - \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\, z_i(\underset{\sim}{y})\, \underset{\sim}{\Phi}(\underset{\sim}{x}_i, \underset{\sim}{y}) = 0$$

$$\Rightarrow \underset{\sim}{w} = \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\, z_i(\underset{\sim}{y})\, \underset{\sim}{\Phi}(\underset{\sim}{x}_i, \underset{\sim}{y}) \longrightarrow (2)$$

$$\frac{\partial L}{\partial \theta_i} = -\sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\, z_i(\underset{\sim}{y}) = 0 \longrightarrow (3)$$

* KKT conditions :

$$\alpha_i(\underset{\sim}{y})\, z_i(\underset{\sim}{y})\left(\underset{\sim}{w}^T \underset{\sim}{\Phi}(\underset{\sim}{x}_i, \underset{\sim}{y}) + \theta_i\right) - 1 + \xi_i = 0 \quad \forall\, i = 1(1)n$$

$$\alpha_i(\underset{\sim}{y}) \geq 0 \quad \forall\, i = 1(1)n \quad \forall\, \underset{\sim}{y} \in \mathcal{Y}$$

Plugging the above results into the Primal Lagrangian,

$$L = \frac{1}{2}\left\|\sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\, z_i(\underset{\sim}{y})\, \underset{\sim}{\Phi}(\underset{\sim}{x}_i, \underset{\sim}{y})\right\|^2 + \frac{C}{2} \sum_{i=1}^{n} \left(\frac{1}{C} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})\right)^2$$

$$- \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) \left(z_i(\underset{\sim}{y}) \Big[\underbrace{\Big\{ \sum_{j=1}^{n} \sum_{\bar{\underset{\sim}{y}}} \alpha_j(\bar{\underset{\sim}{y}})\, \underset{\sim}{\Phi}(\underset{\sim}{x}_j, \bar{\underset{\sim}{y}}) \Big\}}_{w}^T \underset{\sim}{\Phi}(\underset{\sim}{x}_i, \underset{\sim}{y}) + \theta_i \Big] - 1 + \underbrace{\Big[\frac{1}{C} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) \Big]}_{\xi_i} \right)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \sum_{j=1}^{n} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_j(\underset{\sim}{\bar{y}}) z_i(\underset{\sim}{y}) z_j(\underset{\sim}{\bar{y}}) \underbrace{k_{ij}\left((x_i, \underset{\sim}{y}), (x_j, \underset{\sim}{\bar{y}})\right)}_{\text{Hence denoted by } k_{ij}(\underset{\sim}{y}, \underset{\sim}{\bar{y}})}$$

$$+ \frac{1}{2C} \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})$$

$$- \sum_{i} \sum_{\underset{\sim}{y}} \sum_{j} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_j(\underset{\sim}{\bar{y}}) z_i(\underset{\sim}{y}) z_j(\underset{\sim}{\bar{y}}) \overbrace{k_{ij}\left((x_i, \underset{\sim}{y}), (x_j, \underset{\sim}{\bar{y}})\right)}$$

$$- \sum_{i=1}^{n} \theta_i \underbrace{\sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) z_i(\underset{\sim}{y})}_{= 0 \ \ldots \text{from } (3)} + \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})$$

$$- \frac{1}{C} \sum_{i} \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})$$

$$= \frac{-1}{2} \left(\sum_{i, \underset{\sim}{y}} \sum_{j, \underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_j(\underset{\sim}{\bar{y}}) z_i(\underset{\sim}{y}) z_j(\underset{\sim}{\bar{y}}) k_{ij}(\underset{\sim}{y}, \underset{\sim}{\bar{y}}) \right)$$

$$+ \sum_{i=1}^{n} \sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y}) \boxed{- \frac{1}{2C} \sum_{i} \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})}$$

→ maximize w.r.t. α to obtain dual (Eqn 16). Note that $\boxed{}$ term is extra, not present in Eqn 16.

↳ Explained in the next solution.

2.3) Define
$$K^c_{ij}(\underset{\sim}{y},\bar{\underset{\sim}{y}}) = \begin{cases} K_{ii}(\underset{\sim}{y},\bar{\underset{\sim}{y}}) + \frac{1}{c} z_i(\underset{\sim}{y}) z_i(\bar{\underset{\sim}{y}}), & \text{if } i=j, \text{ and,} \\ K_{ij}(\underset{\sim}{y},\bar{\underset{\sim}{y}}), & \text{if } i\neq j \end{cases}$$

then, we can rewrite the dual from 2.2) as follows

$$\underbrace{-\frac{1}{2}\left(\sum_{i,\underset{\sim}{y}}\sum_{j,\bar{\underset{\sim}{y}}} \alpha_i(\underset{\sim}{y})\alpha_j(\bar{\underset{\sim}{y}}) z_i(\underset{\sim}{y}) z_j(\bar{\underset{\sim}{y}}) K^c_{ij}(\underset{\sim}{y},\bar{\underset{\sim}{y}})\right)}_{A} + \sum_i\sum_{\underset{\sim}{y}} \alpha_i(\underset{\sim}{y})$$

if $i=j$,

$$A = -\frac{1}{2}\sum_{i=1}^{n}\sum_{\underset{\sim}{y}}\sum_{\bar{\underset{\sim}{y}}}\left[\alpha_i(\underset{\sim}{y})\alpha_i(\bar{\underset{\sim}{y}}) z_i(\underset{\sim}{y}) z_i(\bar{\underset{\sim}{y}}) \times \left(K_{ii}(\underset{\sim}{y},\bar{\underset{\sim}{y}}) + \frac{1}{c} z_i(\underset{\sim}{y}) z_i(\bar{\underset{\sim}{y}})\right)\right]$$

$$= \underbrace{-\frac{1}{2}\sum_i\sum_{\underset{\sim}{y}}\sum_{\bar{\underset{\sim}{y}}} \alpha_i(\underset{\sim}{y})\alpha_i(\bar{\underset{\sim}{y}}) z_i(\underset{\sim}{y}) z_i(\bar{\underset{\sim}{y}}) K_{ii}(\underset{\sim}{y},\bar{\underset{\sim}{y}})}_{B}$$

$$-\frac{1}{2}\sum_i\sum_{\underset{\sim}{y}}\sum_{\bar{\underset{\sim}{y}}} \alpha_i(\underset{\sim}{y})\alpha_i(\bar{\underset{\sim}{y}}) \underbrace{\left(z_i(\underset{\sim}{y}) z_i(\bar{\underset{\sim}{y}})\right)^2}_{=1} \cdot \frac{1}{C}$$

$$= B \boxed{- \frac{1}{2C} \sum_i \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})}$$

if $i \neq j$,

$$A = -\frac{1}{2} \sum_{i,\underset{\sim}{y}} \sum_{j \neq i, \underset{\sim}{\bar{y}}} \underbrace{\alpha_i(\underset{\sim}{y}) \alpha_j(\underset{\sim}{\bar{y}}) z_i(\underset{\sim}{y}) z_j(\underset{\sim}{\bar{y}}) \cdot K_{ij}(\underset{\sim}{y}, \underset{\sim}{\bar{y}})}_{C}$$

Combining these 2 exhaustive cases ($i=j$ & $i \neq j$), we get

$$A = B \boxed{- \frac{1}{2C} \sum_i \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})} + C$$

$$= -\frac{1}{2} \sum_{i,\underset{\sim}{y}} \sum_{j, \underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_j(\underset{\sim}{\bar{y}}) z_i(\underset{\sim}{y}) z_j(\underset{\sim}{\bar{y}}) K_{ij}(\underset{\sim}{y}, \underset{\sim}{\bar{y}})$$

$$\boxed{- \frac{1}{2C} \sum_i \sum_{\underset{\sim}{y}} \sum_{\underset{\sim}{\bar{y}}} \alpha_i(\underset{\sim}{y}) \alpha_i(\underset{\sim}{\bar{y}})}$$

Thus, we have recovered the missing term in Eq 16 by redefining the kernel function