Assignment 5

Problem1 : Textbook 21.6

In Chapter 21.4, we have derived the mean field method update equation for the Ising model:

$$\mu_i = \tanh\left(\sum_{j \in nbr_i} W_{ij}\mu_j + 0.5(L_i^+ - L_i^-)\right)$$

An alternative way to derive the above update equations is to derive the variational free energy, and to optimize it wrt the a_i explicitly. Complete this derivation.

Recall that the variational free energy $J(q)$ is defined as

$$J(q) = \mathbb{E}_q[log\ q(\mathbf{x})] + \mathbb{E}_q[-log\ \tilde{p}(\mathbf{x})] = -\mathbb{H}(q) + \mathbb{E}_q[E(\mathbf{x})]$$

Problem 2

Suppose we think $p(x|\mathcal{D})\ N(\mu, \sigma^2)$ with $\mu > 0$, and we would like to derive a variational approximation $q(x) \sim \text{Uniform}(0, a)$, with $a > 0$. Derive a.

Problem 3: Textbook 21.7 Forwards vs reverse KL divergence

Consider a factored approximation $q(x, y) = q(x)q(y)$ to a joint distribution $p(x, y)$. Show that to minimize the forwards KL divergence $KL(p||q)$ we should set $q(x) = p(x)$ and $q(y) = p(y)$. That is, the optimal approximation is a product of marginals. Now consider the following joint distribution, where the rows represent y and the columns x.

	x			
	1	2	3	4
1	1/8	1/8	0	0
2	1/8	1/8	0	0
3	0	0	1/4	0
4	0	0	0	1/4

Show that the reverse $KL(q||p)$ for this p has three distinct minima. Identify those minima and evaluate $KL(q||p)$ at each of them. What is the value of $KL(q||p)$ if we set $q(x, y) = p(x)p(y)$?

Problem 4

Consider a three-variable factorial hidden Markov model: We have three hidden variables at each time step t, $X_{1,t}, X_{2,t}, X_{3,t}$ and one observed variable Y_t. Y_t depends just on the hidden variables at its time step $(X_{1,t}, X_{2,t}, X_{3,t})$. $X_{i,t}$ depends on just $X_{i,t-1}$. Draw a Bayes net, factor graph and Markov random field that represent this model for $t = 1 \ldots 4$.

Problem 5

Define the weighted Euclidean distance as

$$d_w(x_i, x_j) = \frac{\sum k = 1^K w_k (x_{ik} - x_{jk})^2}{\sum_{k=1}^{K} w_k}. \tag{1}$$

Show that weighted Euclidean distance on x is equivalent to unweighted Euclidean distance on z, where

$$z_{ik} = x_{ik} \sqrt{\frac{w_k}{\sum_{k=1}^{K} w_k}}. \tag{2}$$

Problem 6: An EM algorithm for a Mixture of Bernoullis

In this problem, you will derive an expectation-maximization (EM) algorithm to cluster black and white images. The inputs $x^{(i)}$ can be thought of as vectors of binary values corresponding to black and white pixel values, and the goal is to cluster the images into groups. You will be using a mixture of Bernoullis model to tackle this problem.

For the sake of brevity, you do not need to substitute in previously derived expression in later problems.For example, beyond question 1.(a) you may use $P(x^{(i)}|p^{(k)})$ in your answers.

1. Mixture of Bernoullis

   (a) Consider a vector of binary random variables, $x \in \{0, 1\}^D$. Assume each variable x_d is drawn from a Bernoulli(p_d) distribution, so $P(x_d = 1) = p_d$. Let $p \in (0, 1)^D$ be the resulting vector of Bernoulli parameters. Write an expression for $P(x|p)$.

   (b) Now suppose we have a mixture of K Bernoulli distributions: each vector $x^{(i)}$ is drawn from some vector of Bernoulli random variables with parameters $p^{(k)}$) , we will call this Bernoulli($p^{(k)}$). Let $\{p^{(1)}, ..., p^{(k)}\} = \mathbf{p}$. Assume a distribution $\pi(k)$ over the selection of which set of Bernoulli parameters $p^{(k)}$ is chosen. Write an expression for $P(x^{(i)}|\mathbf{p}, \pi)$.

   (c) Finally, suppose we have inputs $X = \{x^{(i)}\}_{i=1...n}$. Using the above, write an expression for the log likelihood of the data X, log $P(X|\pi, \mathbf{p})$

2. Expectation step

Assignment 5

(a) Now, we introduce the latent variables for the EM algorithm. Let $z^{(i)} \in \{0,1\}^K$ be an indicator vector, such that $z^{(i)} = 1$ if $x^{(i)}$ was drawn from a Bernoulli($p^{(k)}$), and 0 otherwise. Let $Z = \{z^{(i)}\}_{i=1\ldots n}$.
What is $P(z^{(i)}|\pi)$? What is $P(x^{(i)}|z^{(i)}, \mathbf{p}, \pi)$?

(b) Using the above two quantities, derive the likelihood of the data and the latent variables, $P(Z, X|\pi, \mathbf{p})$.

(c) Let $\eta(z_k^{(i)}) = E[z_k^{(i)}|x^{(i)}, \pi, \mathbf{p}]$. Show that

$$\eta(z_k^{(i)}) = \frac{\pi_k \prod_{d=1}^{D} (p_d^{(k)})^{x_d^{(i)}} (1 - p_d^{(k)})^{1-x_d^{(i)}}}{\sum_j \pi_j \prod_{d=1}^{D} (p_d^{(j)})^{x_d^{(i)}} (1 - p_d^{(j)})^{1-x_d^{(i)}}}$$

Let $\tilde{\mathbf{p}}, \tilde{\pi}$ be the new parameters that we'd like to maximize, so $\mathbf{p}, \pi$ are from the previous iteration. Use this to derive the following final expression for the E step in the expectation-maximization algorithm:

$$E[log\ P(Z, X|\tilde{\mathbf{p}}, \tilde{\pi})|X, \mathbf{p}, \pi)] =$$
$$\sum_{i=1}^{N} \sum_{k=1}^{K} \eta(z_k^{(i)}) \left[log\ \tilde{\pi}_k + \sum_{d=1}^{D} \left(x_d^{(i)} log\ \tilde{p_d^{(k)}} + (1 - x_d^{(i)}) log(1 - \tilde{p_d^{(k)}}) \right) \right]$$

3. Maximization step

(a) We need to maximize the above expression with respect to $\tilde{\mathbf{p}}, \tilde{\pi}$. First, show that the value of $\tilde{\mathbf{p}}$ that maximizes the E step is

$$\tilde{p^{(k)}} = \frac{\sum_{i=1}^{N} \eta(z_k^{(i)}) x^{(i)}}{N_k}$$

where $N_k = \sum_{i=1}^{N} \eta(z_k^{(i)})$.

(b) Show that the value of $\tilde{\pi}$ that maximizes the E step is

$$\tilde{\pi_k} = \frac{N_k}{\sum_{k'} N_{k'}}$$