ES 201 Homework 6, Due 04/27/2017 (in class)

Reading: Chs. 15, 18, Lecs. 21, 22.

Problem 1

The *Iris* data set consists of 150 vectors of features from three species of the *Iris* flower (Iris setosa, Iris virginica and Iris versicolor). The number of feature vectors associated with each species is 50. Let us denote the data set by $\{\mathbf{x}_i\}_{i=1}^n, n = 150$. Each $\mathbf{x}_i \in \mathbb{R}^4, i = 1, \cdots, n$ is associated with one and only one of the three species and corresponds to four features, namely the length and width of the sepals in centimeters (Figure 1).

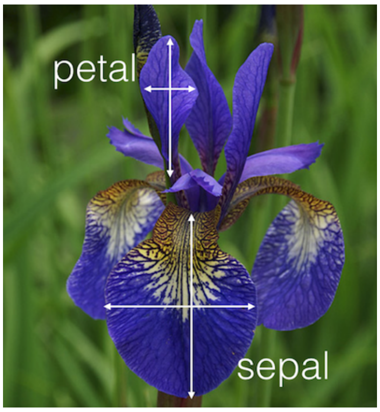

Figure 1: Features collected from each of the 50 flowers in the Iris data set.

We will consider the subset of the data set that consists of the Iris setosa (class label -1) and Iris versicolor (class label $+1$) species, for a total of 100 features vectors. These data are linearly separable in the sepal length/petal length space.

The Jupyter notebook `Iris.ipynb` downloads the dataset from the internet and extracts the relevant subset of the data.

(a) Implement the *online* Perceptron solution to the problem of separating Iris-setosa and Iris-versicolor in the sepal length/petal length space. Use a learning rate $\mu = 0.1$ and 10 epochs (i.e. cycle through the data 10 times). Plot the Perceptron decision boundary in the same figure as the original data set, color coded by class.

(b) Plot the error rate (fraction of misclassified examples) as a function of iteration. Vary the learning rate and comment on what you observe.

(c) Plot the Perceptron decision boundary and the decision boundary from the maximum-margin classifier obtained with an SVM (you may use `scikit-learn` or any package of your choice). Compare the two decision boundaries and comment.

Problem 2

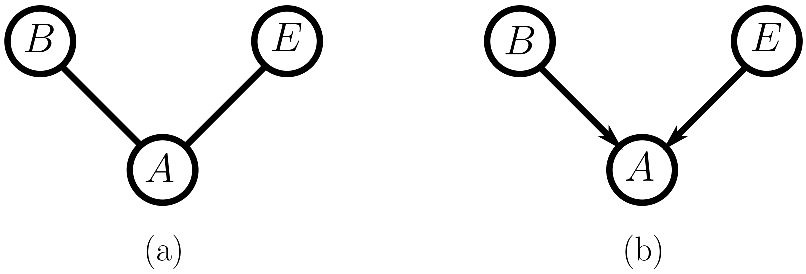

Figure 2: Two graphical models of the triggering of an alarm system ($A \in \{0,1\}$), by one of two events: a burglary ($B \in \{0,1\}$) and/or an earthquake $E \in \{0,1\}$. (a) Markov Random Field (MRF) model; (b) Bayesian Network model (BN).

Figure 2 shows two graphical models of three binary events A, B and E. The event $A \in \{0,1\}$ represents the triggering of an alarm system, while the events $B \in \{0,1\}$ and $E \in \{0,1\}$ represent the occurrence of a burglary and an earthquake respectively. The BN explicitly represents the *causal* relationship between these three events, while the MRF does not.

(a) Write down the joint probability mass function (pmf) $P[A, E, B]$ under the MRF in terms of potentials $\Psi_1(A, B)$ and $\Psi_2(A, E)$ and a normalization constant.

(b) Under the MRF, what does $P[B, E|A]$ equal? Explain why the expression you obtain is unreasonable given the inherent causal nature of the process being modelled.

(c) Write down the joint pmf $P[A, E, B]$ under the BN.

(d) Consider the table below that represents $P[A|E, B]$. Fill in values for the rest of the entries.

Table 1: *Conditional* pmf $P[A|E, B]$

A \ (E, B)	00	01	10	11
0			0	
1		1		

(e) Suppose $P[B = 1] = \delta_B$ and $P[E = 1] = \delta_E, \delta_B$. Assume $\delta_B << 1$ and $\delta_E << 1$, i.e. burglaries and earthquakes are events with rare probabilities. Fill in the table below that represents the *joint* pmf $P[A, E, B]$.

Table 2: *Joint* pmf $P[A, E, B]$

A \ (E, B)	00	01	10	11
0				
1				

(f) Compute $P[B = 1|E = 1, A = 1]$ and $P[B = 1|E = 0, A = 1]$. Explain in words what you can conclude by comparing these quantities? Is your conclusion reasonable given the causal nature of the process being modelled? What can you conclude regarding the power of the BN compared to the MRF in this example?