

Problem 3 Ising graphical Model

1. Ising Markov random fields:

Recall from lecture that an Ising model is a model where the nodes have binary variables and the distribution has pairwise interactions. Let $\mathbf{X} = \{X_1, \dots, X_p\}$ be binary random variables associated with an undirected graph. For a face detection task in image processing, each variable in \mathbf{X} might represent a labeling of the corresponding pixel location as “face” or “not face”. An Ising Markov random field takes the form:

$$p(\mathbf{X}; \theta) \propto \exp \left\{ \sum_{s \in V} \theta_s X_s + \sum_{(s,t) \in E} \theta_{s,t} X_s X_t \right\} \quad (6)$$

For a given node $i \in V$, show that the conditional $P(X_i = 1 | \mathbf{X}_{-i}; \theta)$ is given by a logistic regression model.

[Solution]: First we show

$$\frac{P(X_i = 1, \mathbf{X}_{-i}; \theta)}{P(X_i = 0, \mathbf{X}_{-i}; \theta)} = \exp \left\{ 2 + 2 \sum_{s \neq i} \theta_{s,i} X_s \right\}.$$

We then plug in:

$$P(X_i = 1 | \mathbf{X}_{-i}; \theta) = \frac{P(X_i = 1, \mathbf{X}_{-i})}{P(X_i = 1, \mathbf{X}_{-i}) + P(X_i = 0, \mathbf{X}_{-i})} = \frac{\frac{P(X_i=1, \mathbf{X}_{-i})}{P(X_i=0, \mathbf{X}_{-i})}}{\frac{P(X_i=1, \mathbf{X}_{-i})}{P(X_i=0, \mathbf{X}_{-i})} + 1} = \frac{\exp\{2 + 2 \sum_{s \neq i} \theta_{s,i} X_s\}}{\exp\{2 + 2 \sum_{s \neq i} \theta_{s,i} X_s\} + 1}.$$

Note that the zeros in the logistic regression coefficients correspond to the zeros in the graphical model. This motivates the use of neighborhood selection to learn the graph structure of Ising models.

2. Ising conditional random fields:

In this question, we consider a conditional random field (CRF) extension of the Ising model. We are now able to condition on feature variables $\mathbf{W} = \{W_1, \dots, W_q\}$. Going back to the face detection example, each variable in \mathbf{X} corresponds to whether a pixel location is a face or not, while variables in \mathbf{W} might be observed image caption information. The conditional distribution is written as:

$$p(\mathbf{X} | \mathbf{W}; \theta, \beta) = \exp \left\{ \sum_{s \in V} \theta_s X_s + \sum_{(s,t) \in E} \theta_{s,t} X_s X_t + \sum_{s \in V, u \in [q]} \beta_{su} X_s W_u \right\} / Z(\mathbf{W}, \theta, \beta) \quad (7)$$

$$Z(\mathbf{W}, \theta, \beta) = \sum_{\mathbf{X}} \exp \left\{ \sum_{s \in V} \theta_s X_s + \sum_{(s,t) \in E} \theta_{st} X_s X_t + \sum_{s \in V, u \in [q]} \beta_{su} X_s W_u \right\} \quad (8)$$

Note that we now have a parameter β relating \mathbf{X} to \mathbf{W} . Often to learn CRFs we minimize the log-likelihood using gradient descent. Compute the derivative of the log-likelihood on a dataset containing a single sample with respect to each of the parameters (θ_s, θ_{st} , and β_{su}).

[Solution]: The conditional log-likelihood is

$$\sum_{s \in V} \theta_s X_s + \sum_{(s,t) \in E} \theta_{st} X_s X_t + \sum_{s \in V, u} \beta_{su} X_s W_u - \log Z(\mathbf{W}, \theta, \beta).$$

So the derivatives are:

$$\begin{aligned}
\frac{\partial}{\partial \theta_s} [\log p(\mathbf{X}|\mathbf{W}; \theta, \beta)] &= X_s - \frac{\sum_{\mathbf{X}'} X'_s \exp \left\{ \sum_{s \in V} \theta_s X'_s + \sum_{(s,t) \in E} \theta_{st} X'_s X'_t + \sum_{s \in V, u} \beta_{su} X'_s W_u \right\}}{Z(\mathbf{W}, \theta, \beta)} \\
&= X_s - \sum_{\mathbf{X}'} X'_s P(\mathbf{X}'|\mathbf{W}; \theta, \beta) \\
&= X_s - E_{P(\mathbf{X}'|\mathbf{W}; \theta, \beta)}[X'_s] \\
\frac{\partial}{\partial \theta_{st}} [\log p(\mathbf{X}|\mathbf{W}; \theta, \beta)] &= X_s X_t - \frac{\sum_{\mathbf{X}'} X'_s X'_t \exp \left\{ \sum_{s \in V} \theta_s X'_s + \sum_{(s,t) \in E} \theta_{st} X'_s X'_t + \sum_{s \in V, u} \beta_{su} X'_s W_u \right\}}{Z(\mathbf{W}, \theta, \beta)} \\
&= X_s X_t - \sum_{\mathbf{X}'} X'_s X'_t P(\mathbf{X}'|\mathbf{W}; \theta, \beta) \\
&= X_s X_t - E_{P(\mathbf{X}'|\mathbf{W}; \theta, \beta)}[X'_s X'_t] \\
\frac{\partial}{\partial \beta_{su}} [\log p(\mathbf{X}|\mathbf{W}; \theta, \beta)] &= W_u X_s - \frac{\sum_{\mathbf{X}'} W_u X'_s \exp \left\{ \sum_{s \in V} \theta_s X'_s + \sum_{(s,t) \in E} \theta_{st} X'_s X'_t + \sum_{s \in V, u} \beta_{su} X'_s W_u \right\}}{Z(\mathbf{W}, \theta, \beta)} \\
&= W_u X_s - W_u \sum_{\mathbf{X}'} P(\mathbf{X}'|\mathbf{W}; \theta, \beta) \\
&= W_u X_s - W_u E_{P(\mathbf{X}'|\mathbf{W}; \theta, \beta)}[X'_s]
\end{aligned}$$

Problem 5: The undirected graph in Fig. 1 represents a Markov network with nodes x_1, x_2, x_3, x_4, x_5 counting clockwise around the pentagon with potentials $\phi(x_i, x_j)$. Show that the joint distribution can be written as

$$p(x_1, x_2, x_3, x_4, x_5) = \frac{p(x_1, x_2, x_5)p(x_2, x_4, x_5)p(x_2, x_3, x_4)}{p(x_2, x_5)p(x_2, x_4)}$$

and also express the marginal probability tables (i.e., the terms in the numerator and denominator of the above formula) explicitly as functions of the pair-wise potentials $\phi(x_i, x_j)$.

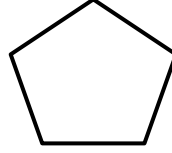


Fig. 1: Markov Network for Exercise 4.6

Answer:

Using the obvious shorthand, we have

$$p_{12345} = \phi_{45}\phi_{12}\phi_{23}\phi_{34}\phi_{51}$$

By summing over 3 and 4, we obtain

$$p_{125} = \phi_{12}\phi_{51} \sum_{34} \phi_{23}\phi_{34} \phi_{45},$$

Similarly, we obtain

$$p_{245} = \phi_{45} \sum_{13} \phi_{12}\phi_{23}\phi_{34}\phi_{51}, \quad p_{234} = \phi_{23}\phi_{34} \sum_{15} \phi_{12}\phi_{45}\phi_{51},$$

Multiplying these we obtain

$$\begin{aligned} p_{125}p_{245}p_{234} \\ = \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51} \left(\sum_{34} \phi_{23}\phi_{34}\phi_{45} \right) \left(\sum_{13} \phi_{12}\phi_{23}\phi_{34}\phi_{51} \right) \left(\sum_{15} \phi_{12}\phi_{45}\phi_{51} \right) \end{aligned}$$

Also,

$$p_{25} = \sum_{134} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}, \quad p_{24} = \sum_{135} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}$$

Hence

$$\begin{aligned}
\frac{p_{125}p_{245}p_{234}}{p_{25}p_{24}} &= p \frac{(\sum_{34} \phi_{23}\phi_{34}\phi_{45}) (\sum_{13} \phi_{12}\phi_{23}\phi_{34}\phi_{51}) (\sum_{15} \phi_{12}\phi_{45}\phi_{51})}{(\sum_{134} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}) (\sum_{135} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51})} \\
&= p \frac{(\sum_{34} \phi_{23}\phi_{34}\phi_{45}) (\sum_{13} \phi_{12}\phi_{23}\phi_{34}\phi_{51}) (\sum_{15} \phi_{12}\phi_{45}\phi_{51})}{(\sum_1 \phi_{12}\phi_{51} \sum_{34} \phi_{23}\phi_{34}\phi_{45}) (\sum_3 \phi_{23}\phi_{34} \sum_{15} \phi_{12}\phi_{45}\phi_{51})} \\
&= p \frac{(\sum_{34} \phi_{23}\phi_{34}\phi_{45}) (\sum_1 \phi_{12}\phi_{51}) (\sum_3 \phi_{23}\phi_{34}) (\sum_{15} \phi_{12}\phi_{45}\phi_{51})}{(\sum_1 \phi_{12}\phi_{51} \sum_{34} \phi_{23}\phi_{34}\phi_{45}) (\sum_3 \phi_{23}\phi_{34} \sum_{15} \phi_{12}\phi_{45}\phi_{51})} \\
&= p
\end{aligned}$$

The reason this works is more clearly seen when considering the junction tree representation (we will study later in Inference Section of the course): by triangulation, we can write the distribution using cliques on $(1, 2, 5)$, $(2, 4, 5)$ and $(2, 3, 4)$, with $(2, 4)$ and $(2, 5)$ being their separators.

4. Hidden Markov Model

4.1 Base Line

Solution: The solution code is in the “hmm_sol” folder.

A Replace infrequent words with _RARE_:

```
python src/replace_infreq.py data/gene.train gene.train.rare
```

B Count frequencies:

```
python src/count_freqs.py gene.train.rare > gene.train.rare.counts
```

C Run baseline decoding:

```
python src/hmm_baseline.py gene.train.rare.counts data/gene.test gene_test.p1.out
```

D Evaluate:

```
python src/eval_gene_tagger.py data/gene.key gene_test.p1.out
```

Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

	precision	recall	F1-Score
GENE:	0.158861	0.660436	0.256116

4.2 HMM with trigram features

Solution: The solution code is in the “hmm_sol” folder.

A Replace infrequent words with _RARE_:

```
python src/replace_infreq.py data/gene.train gene.train.rare
```

B Count frequencies:

```
python src/count_freqs.py gene.train.rare > gene.train.rare.counts
```

C Run Viterbi decoding:

```
python src/hmm_trigram.py gene.train.rare.counts data/gene.test gene_test.p2.out
```

D Evaluate:

```
python src/eval_gene_tagger.py data/gene.key gene_test.p2.out
```

Found 191 GENEs. Expected 642 GENEs; Correct: 104.

	precision	recall	F1-Score
GENE:	0.544503	0.161994	0.249700