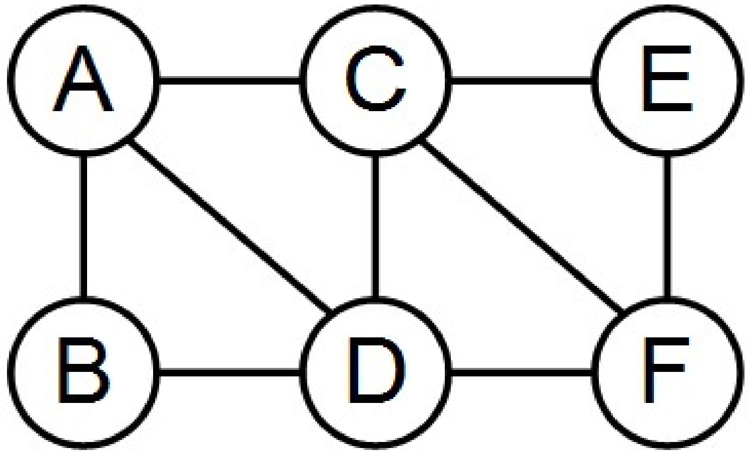
Markov Networks

In the third homework you will review some concepts about Markov Networks that we saw in class, and you will also learn more about some additional topics.



For the theoretical exercises, please be as explicit and clear as possible. Furthermore, use the $\LaTeX$ math mode that notebooks offer.

If further questions arise, please use the slack channel, or write me to esjimenezro@iteso.mx.

Imagen recuperada de: https://upload.wikimedia.org/wikipedia/en/7/7b/A_simple_Markov_network.png.

---

1. Logistic regression as CRFs

Consider a CRF:

- Over the binary RVs $\bar{X} = \{X_1, \ldots, X_n\}$ and $\bar{Y} = \{Y\}$.
- There are pairwise edges between Y and each X_i.
- The factors are defined as:

$$\phi_i(Y, X_i) = \exp(w_i \mathbf{1}\{X_i = 1, Y = 1\}),$$

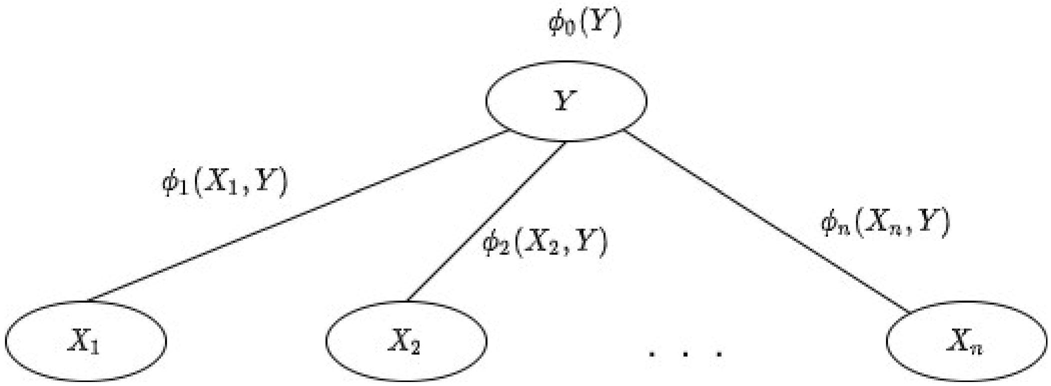
  where $w_i \in \mathbb{R}$ and $\mathbf{1}$ stands for the indicator function.
- Moreover, there is a single-node factor $\phi_0(Y) = \exp(w_0 \mathbf{1}\{Y = 1\})$.

Show that the conditional probability distribution this CRF encodes corresponds to the logistic regression distribution:

$$P(Y = 1 | \bar{x}) = \frac{\exp\left(w_0 + \sum_{i=1}^n w_i x_i\right)}{1 + \exp\left(w_0 + \sum_{i=1}^n w_i x_i\right)}$$

```python
from IPython.display import Image
```

```
Image("figures/LogisticCRF.png")
```



Definition. A **conditional random field (CRF)** is an undirected graph $\mathcal{H}$ whose nodes correspond to $\bar{X} \cup \bar{Y}$, parameterized by a set of factors $\Phi = \{\phi_1(\bar{D}_1), \ldots, \phi_k(\bar{D}_k)\}$, such that $\bar{D}_i \not\subset \bar{X}$. This network encodes a conditional probability distribution as follows

$$P(\bar{Y}|\bar{X}) = \frac{1}{Z(\bar{X})}\tilde{P}(\bar{Y}, \bar{X}),$$

$$\tilde{P}(\bar{Y}, \bar{X}) = \prod_{i=1}^{k} \phi_i(\bar{D}_i),$$

$$Z(\bar{X}) = \sum_{\bar{Y}} \tilde{P}(\bar{Y}, \bar{X}).$$

Prueba:

$$\tilde{P}(\bar{Y}, \bar{X}) = \phi_0(Y)\prod_{i=1}^{n} \phi_1(Y, X_i)$$

si $Y = 1$

$$\phi_0(Y = 1) = \exp(w_0)$$

$$\phi_1(Y = 1, X_i = 1) = \exp(w_i X_i)$$

$$\tilde{P}(\bar{Y} = 1, \bar{X}) = \exp(w_0 + \sum_{i=1}^{n} w_i X_i)$$

Si

$$Z(X) = \sum_{Y} \tilde{P}(\bar{Y}, \bar{X})$$

entonces:

$$Z(X) = \sum_{Y} \exp(w_0 + \sum_{i=1}^{n} w_i X_i) = 1 + \exp(w_0 + \sum_{i=1}^{n} w_i X_i)$$

Por lo tanto:

$$P(Y = 1 | \bar{X}) = \frac{\exp(w_0 + \sum_{i=1}^{n} w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^{n} w_i X_i)}$$

2. Multinomial logistic as CRFs

We will extend the result above to multinomial models.

- Assume now that $\bar{X} = \{X_1, \ldots, X_n\}$ and $\bar{Y} = \{Y\}$ are multivalued.
- The factors are defined as:

$$\phi_i(y^m, x_i^l) = \exp(w_i \mathbf{1}\{X_i = x_i^l, Y = y^m\}),$$

  where $w_i \in \mathrm{R}$ and $\mathbf{1}$ stands for the indicator function.
- Again, there is a single-node factor $\phi_0(Y) = \exp(w_0 \mathbf{1}\{Y = y^m\})$.

Show that the conditional distribution $P(Y|X_1, \ldots, X_k)$ defined by a CRF with the above factors is equivalent to the [multinomial logistic regression distribution (https://en.wikipedia.org/wiki/Multinomial_logistic_regression#As_a_set_of_independent_binary_regressions)](https://en.wikipedia.org/wiki/Multinomial_logistic_regression#As_a_set_of_independent_binary_regressions).

Demostración 2

$\phi_i^{ml}(Y, X_i) = \exp(w_i^{ml}\mathbf{1}\{Y = y^m, X = xi^l\})$ $\phi_0^m(Y) = \exp(w_0^m\mathbf{1}\{Y = y^m\})$ Donde:

- $i = 1, \ldots, n$ - given X RVs
- $l = 1, \ldots, ni$
- $m = 1, \ldots, m$ - given Y $\bar{P}(Y|\bar{X}) = \prod_{m=1}^{M} phi_0^m(Y) \prod_{i=1}^{n} \prod_{l=1}^{ni} phi_i^{ml}(Y, Xi)$

$P(Y|\bar{X}) = \prod_{l=1}^{ni} phi_i^{ml}(Y, Xi)$

$P(Y|\bar{X}) = \frac{1}{Z(\bar{X})}\bar{P}(Y, \bar{X})$

$Z(\bar{X}) = \sum_Y \bar{P}(Y, \bar{X})$

$Y = y^j$

$\phi_0^j(Y = y^j) = exp(w_0^j)$

$\phi_0^k(Y = y^j) = 1; k \neq j$

$\phi_i^{jl}(Y = y^j, Xi = xi) = exp(w_i^{jl}\mathbf{1}\{Xi = xi^l\}) \ = exp(w_i^{jl}x_i l)$

$\phi_i^{kl}(Y = y^k, X_i = x_i) = 1$

$\bar{P}(Y = y^j|\bar{X}) = exp(w_0^j + \sum_{i=1}^{n}\sum_{l=1}^{ni} w_i^{jl}X_i^l)$

$Z(\bar{X}) = \sum_Y \bar{P}(Y, \bar{X})$

$\sum_{j=i}^{M} exp(w_0^j + \sum_{i=1}^{n}\sum_{l=1}^{ni} w_i^{jl}x_i^l)$

$P(Y = y^j|\bar{X}) = \frac{\bar{P}(Y=y^j|\bar{X})}{Z(\bar{X})}$

3. Log-linear model

Can you rewrite the above CRF as a log-linear model?

$$\frac{exp(w_0^j + \sum_{i=1}^{n} \sum_{l=1}^{ni} w_i^{jl} X_i^l)}{\sum_{j=i}^{M} exp(w_0^j + \sum_{i=1}^{n} \sum_{l=1}^{ni} w_i^{jl} x_i^l)}$$

1. What are the form of the features?

$$X_i^l$$

2. What are the parameters?

$$w_0^M, w_0^j, w_i^{jl}$$

3. What is the final CPD form as a log-linear model?

$$P(Y|\bar{X}) = \frac{exp(w_0^j + \sum_{i=1}^{n} \sum_{l=1}^{ni} w_i^{jl} X_i^l)}{\sum_{j=i}^{M} exp(w_0^j + \sum_{i=1}^{n} \sum_{l=1}^{ni} w_i^{jl} x_i^l)}$$

4. Metric Markov random fields

Read box 4.D (pages 127-128) and answer:

1. Which are the trhee properties that a metric must satisfy?
2. Give, at least, five examples of metric functions.
3. What is a possible application of Metric MRFs?

1. Three properties that a metric must satisfy

> 1. **Reflexivity:** $\mu(vk, vl) = 0$ if and only if $k = l$;
> 2. **Symmetry:** $\mu(vk, vl) = \mu(vl, vk)$;
> 3. **Triangle Inequality:** $\mu(vk, vl) + \mu(vl, vm) \geq \mu(vk, vm)$.
>
> **2. Five examples of metric functions** Truncated p-norm
>
> **3. Possible application of MRFs** R= Computer Vision

Created with Jupyter by Esteban Jiménez Rodríguez.