Probabilistic Graphical Models

VSC

Universidade do Porto

2020

Outline

Introduction

Probabilistic Graphical Models

Introduction

- Regression targets continuous outputs
- In computing, odten we look at discrete data
- Ofcourse, we can keep on using regression:

$$\sum a_i X_i \geq \theta$$

- Logistic Regression adapts LR to the discrete world.:

Probabilistic Graphical Models

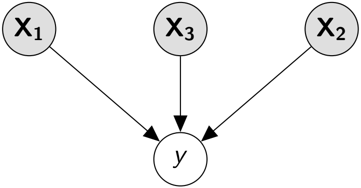
- World is a collection of random variantes $X_1, \ldots X_N$ with joint $\rho(X_1 \ldots X_N)$
- Learn the distribution from data
- Compute conditional probability $Pr(Y|X \ldots)$

Challenges

- Best Model:
  - Directed
  - Undirected
- Learn the distribution from data
  - Compute probability $\Pi Pr(X \ldots)$
  - Compute conditional probability $\Pi Pr(Y_i|X \ldots)$
- Answer Queries?
  - Exact or Approximate?
  - Max or Total?

Intuition

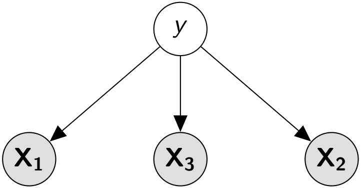
- Imagine we are interested in Y and have evidence X_1, X_2, X_N.



- This is model is perfect,
- but we have to know all $Pr(X|Y)$

Naive Bayes Model

- Bayes rule: $Pr(Y|X\ldots) = \frac{Pr(X...|Y)Pr(Y)}{Pr(X...)}$
- $Pr(Y)$ is the frequency of Y in data set;
- $Pr(X\ldots)$ nobody cares (why?), so
- Simplify
  $P(X_1 \ldots X_N|Y) \rightarrow Pr(X_1|X_N)\ldots Pr(X_N|Y)$



Naive Bayes: the classifier?

- $\frac{Pr(Y|X_1 ... X_N)}{Pr(\overline{Y}|X_1 ... X_N)} > 1$: positive
- $\frac{\frac{Pr(X_1|Y)...Pr(X_N|Y)Pr(Y)}{Pr(X_1...X_N)}}{\frac{Pr(X_1|\overline{Y})...Pr(X_N|\overline{Y})Pr(\overline{Y})}{Pr(X_1...X_N)}} = \frac{Pr(X_1|Y)...Pr(X_N|Y)Pr(Y)}{Pr(X_1|\overline{Y})...Pr(X_N|\overline{Y})Pr(\overline{Y})}$
- $= \frac{Pr(X_1|Y)}{Pr(X_N|\overline{Y})} \times \ldots \times \frac{Pr(X_1|Y)}{Pr(X_N|\overline{Y})} \times \frac{Pr(Y)}{Pr(\overline{Y})}$
- Taking the logs
  $\sum_{X_1}^{X_N} \log Pr(X_i|Y) - \log Pr(X_i|\overline{Y}) +$
  $\log Pr(Y) - \log Pr(\overline{Y}) > 0$

Parameters

- Usually classifier optimises $\Pi Pr(Y|X_1 \ldots X_N)$
- BNs often maximise $\Pi Pr(Y, X_1 \ldots X_N|M)$, the best model forall data,
- More precisely, assume exaamples are iid and use lns': $LogL = \sum_D \log Pr(x_1 \ldots x_N, y)$
- in NB this gives: $LogL = \sum_D \sum_{x_1}^{x_n} \log Pr(x_i|y) + \log Pr(y)$
- to choose parameters, maximize LogL.

Maths

- If $\mathbf{X}$, $\mathbf{X}^{max}$ is maxarg $F(\mathbf{x})$ if $\nabla F(\mathbf{x}) = 0$.
- $\frac{df(g(x))}{dx} = f'(g(x))g'(x)$
- $\frac{df(y)}{dx} = 0$ (if x does not occur in $f(y)$)
- $\log'(x) = \frac{1}{x}$, hence $\log'(f(x)) = \frac{f'(x)}{f(x)}$

Estimating $Pr(x_i|y)$

- We assume binary RVs;
- P positive, N neg;
- X_i is true TP times when y is true, and false FN, so $TP + FN \leq P$

Estimating $Pr(x_i|y)$

- $\frac{\partial LogL}{\partial Pr(x_i|y)} = 0$
- $TP\frac{\partial \ln Pr(x_i|y)}{\partial Pr(x_i|y)} + (P - TP)\frac{\partial \ln(1-Pr(x_i|y))}{\partial Pr(x_i|y)} = 0$
- $\frac{TP}{Pr(x_i|y)} - \frac{P-TP}{1-Pr(x_i|y)} = 0$
- $(P - TP)Pr(x_i|y) = TP(1 - Pr(x_i|y))$