Обучение байесовских сетей. Алгоритм EM.

Сергей Николенко

Машинное обучение — ИТМО, осень 2006

Outline

Суть лекции

- На предыдущих лекциях мы поняли, что такое байесовские сети доверия.
- Теперь нужно научиться их обучать, то есть из набора тестовых данных получать вероятности.

Варианты

- Структура сети:
  - дана заранее;
  - нужно вывести из тестовых примеров.
- Каждый тестовый пример:
  - содержит значения всех переменных сети;
  - содержит некоторые переменные, а о некоторых ничего не известно.

Простейший случай

Простейший случай — когда структура сети дана заранее, и каждый тестовый пример содержит значения всех переменных. Тогда можно просто получить оценки условных вероятностей, подсчитав частоты встречаемости тех или иных переменных.

Второй случай

Уже не настолько простой случай — когда структура сети дана, но не во всех примерах заданы все переменные сети. Здесь уже не обойтись подсчётом частот.

Градиентный подъем

- Будем использовать метод градиентного подъёма.
- Пространство гипотез — возможные элементы таблиц условных вероятностей.
- Задача — максимизировать $p(D|h)$ — условную вероятность имеющихся тестовых данных при условии различных гипотез.
- Метод — следовать градиенту $\ln p(D|h)$ по переменным, соответствующим элементам таблиц условных вероятностей.

Обозначения

- Узлы сети обозначим через x_i, их значения — через v_{ij} (у каждого узла своё множество значений).
- w_{ijk} — это условная вероятность того, что x_i примет значение v_{ij} при условии, что его родители $\text{pa}(x_i)$ примут значение, заданное u_{ik}; т.е. u_{ik} может выражать сразу несколько значений переменных.

Вывод формулы

- Предполагаем, что тестовые примеры независимы:

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} = \frac{\partial}{\partial w_{ijk}} \ln \prod_{d \in D} p(d|h) = \sum_{d \in D} \frac{1}{p(d|h)} \frac{\partial p(d|h)}{\partial w_{ijk}}.$$

- Теперь перейдём к сумме по значениям родителей:

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} = \sum_{d \in D} \frac{1}{p(d|h)} \times$$

$$\times \frac{\partial}{\partial w_{ijk}} \sum_{j',k'} p(d|x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}, h) p(x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}|h)$$

$$= \sum_{d \in D} \frac{1}{p(d|h)} \frac{\partial}{\partial w_{ijk}} \sum_{j',k'} p(d|x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}, h) \times$$

$$\times p(x_i = v_{ij'}|\mathrm{pa}(x_i) = u_{ik'}, h) p(\mathrm{pa}(x_i) = u_{ik'}|h).$$

Вывод формулы

- Теперь перейдём к сумме по значениям родителей:

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} = \sum_{d \in D} \frac{1}{p(d|h)} \times$$

$$\times \frac{\partial}{\partial w_{ijk}} \sum_{j',k'} p(d|x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}, h) p(x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}|h)$$

$$= \sum_{d \in D} \frac{1}{p(d|h)} \frac{\partial}{\partial w_{ijk}} \sum_{j',k'} p(d|x_i = v_{ij'}, \mathrm{pa}(x_i) = u_{ik'}, h) \times$$

$$\times p(x_i = v_{ij'}|\mathrm{pa}(x_i) = u_{ik'}, h) p(\mathrm{pa}(x_i) = u_{ik'}|h).$$

- $w_{ijk} = p(x_i = v_{ij'}|\mathrm{pa}(x_i) = u_{ik'}, h)$, поэтому в сумме только один ненулевой член: когда $j' = j$ и $i' = i$.

Вывод формулы

- $w_{ijk} = p(x_i = v_{ij'} | \mathrm{pa}(x_i) = u_{ik'}, h)$, поэтому в сумме только один ненулевой член: когда $j' = j$ и $i' = i$.

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} = \sum_{d \in D} \frac{1}{p(d|h)} \frac{\partial}{\partial w_{ijk}} p(d|v_{ij}, u_{ik}, h) w_{ijk} p(u_{ik}|h) =$$

$$= \sum_{d \in D} \frac{1}{p(d|h)} \frac{\partial}{\partial w_{ijk}} p(d|v_{ij}, u_{ik}, h) p(u_{ik}|h).$$

Вывод формулы

- Перепишем это по теореме Байеса:

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} = \sum_{d \in D} \frac{1}{p(d|h)} \frac{p(v_{ij}, u_{ik}|d,h)p(d|h)p(u_{ik}|h)}{p(v_{ij}, u_{ik}|h)} =$$

$$= \sum_{d \in D} \frac{p(v_{ij}, u_{ik}|d,h)p(u_{ik}|h)}{p(v_{ij}, u_{ik}|h)} =$$

$$= \sum_{d \in D} \frac{p(v_{ij}, u_{ik}|d,h)}{p(v_{ij}|u_{ik},h)} =$$

$$= \sum_{d \in D} \frac{p(v_{ij}, u_{ik}|d,h)}{w_{ijk}}.$$

Вероятности

$$\frac{\partial \ln p(D|h)}{\partial w_{ijk}} \sum_{d \in D} \frac{p(v_{ij}, u_{ik}|d, h)}{w_{ijk}}.$$

Вероятности $p(v_{ij}, u_{ik}|d, h)$ уже можно оценить из исходных данных. Либо x_i и $\text{pa}(x_i)$ уже есть среди наблюдаемых величин, либо их можно вывести посредством стандартного байесовского вывода из наблюдаемых величин.

Алгоритм

- Проинициализировать w_{ijk} случайными значениями так, чтобы $w_{ijk} \in [0,1]$ и $\forall i,k \sum_j w_{ijk} = 1$.
- Для каждого тестового примера $d \in D$:
  - Для каждого веса w_{ijk}:

$$w_{ijk} \longleftarrow w_{ijk} + \eta \sum_{d \in D} \frac{p(v_{ij}, u_{ik} | d, h)}{w_{ijk}}.$$

Упражнения

Упражнение

Добавить возможность обучения по методу градиентного подъёма в имеющуюся реализацию байесовских сетей доверия.

Outline

Постановка задачи

- Часто возникает ситуация, когда в имеющихся данных некоторые переменные присутствуют, а некоторые — отсутствуют.
- Например, в нашем примере байесовской сети может сложиться ситуация, когда известно, что компьютер упал со стола, и винчестер не читается, но про дефектные секторы ничего наверняка не известно.

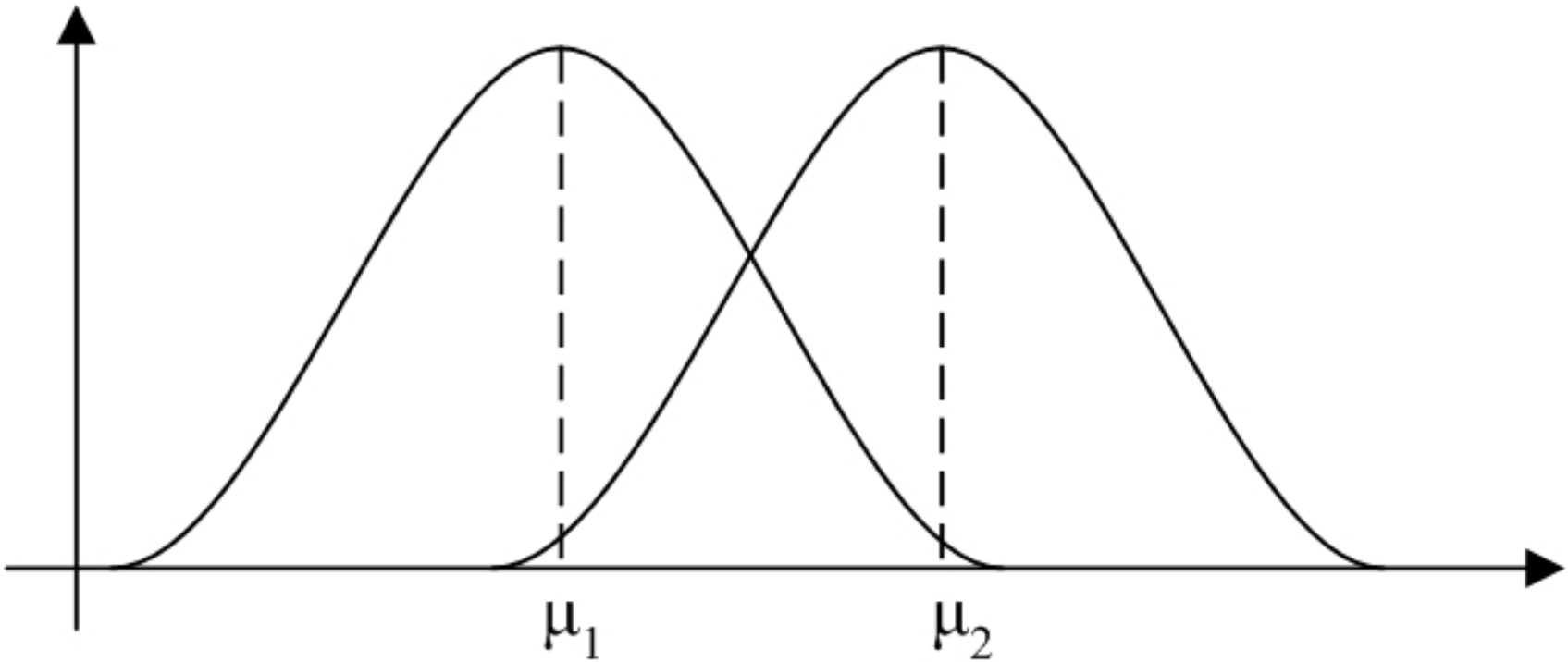
Постановка задачи

- Даны результаты сэмплирования распределения вероятностей с несколькими параметрами, из которых известны не все.
- Эти неизвестные параметры тоже расцениваются как случайные величины.
- Задача — найти наиболее вероятную гипотезу, то есть ту гипотезу h, которая максимизирует

$$E[\ln p(D|h).$$

Частный случай

Построим один из простейших примеров применения алгоритма EM. Пусть случайная переменная x сэмплируется из суммы двух нормальных распределений. Дисперсии даны (одинаковые), нужно найти только средние μ_1, μ_2.



Одно распределение

Если бы было одно распределение, всё было бы как раньше:

$$\mu_{ML} = \operatorname{argmin}_{\mu} \sum_{i=1}^{m} (x_i - \mu)^2,$$

и ответом было бы среднее арифметическое

$$\mu_{ML} = \frac{1}{m} \sum_{i=1}^{m} x_i.$$

Два распределения

- Теперь нельзя понять, какие x_i были порождены каким распределением.
- Это классический пример *скрытых переменных*.

Два распределения

- Как описать один тестовый пример полностью?

Два распределения

- Как описать один тестовый пример полностью?
- Например, как тройку $\langle x_i, z_{i1}, z_{i2} \rangle$, где $z_{ij} = 1$ iff x_i был сгенерирован j-м распределением.

Суть алгоритма EM

- Сгенерировать какую-нибудь гипотезу $h = (\mu_1, \mu_2)$.
- Пока не дойдём до локального максимума:
  - Вычислить ожидание $E(z_{ij})$ в предположении текущей гипотезы.
  - Вычислить новую гипотезу $h' = (\mu_1', \mu_2')$, предполагая, что z_{ij} принимают значения $E(z_{ij})$.

В нашем примере

В нашем примере:

$$E(z_{ij}) = \frac{p(x = x_i|\mu = \mu_j)}{p(x = x_i|\mu = \mu_1) + p(x = x_i|\mu = \mu_2)} = \\ = \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{e^{-\frac{1}{2\sigma^2}(x_i - \mu_1)^2} + e^{-\frac{1}{2\sigma^2}(x_i - \mu_2)^2}}.$$

Мы подсчитываем эти ожидания, а потом подправляем гипотезу:

$$\mu_j \leftarrow \frac{1}{m}\sum_{i=1}^{m} E(z_{ij})x_i.$$

Данные

- Итак, дан набор тестовых примеров $X = \{x_1, \ldots, x_m\}$, имеется также набор скрытых переменных $Z = \{z_1, \ldots, z_m\}$ для этих примеров, зависящих от параметров θ, и полный набор данных $Y = X \cup Z$.
- EM ищет гипотезу максимального правдоподобия h', максимизируя $E(\ln p(Y|h'))$. Для этого он использует уже имеющуюся гипотезу для оценки скрытых параметров.
- Введём функцию $Q(h'|h)$ для оценки ожидания как функции от h':

$$Q(h'|h) = E(\ln p(Y|h')|h, X).$$

Алгоритм EM

- Повторять до тех пор, пока не выполнено условие сходимости:
  - Estimation:
$$Q(h'|h) \leftarrow E(\ln p(Y|h')|h, X).$$
  - Maximization:
$$h \leftarrow \operatorname{argmax}_{h'} Q(h'|h).$$

Упражнения

Упражнение

Разработать версию алгоритма EM, оценивающую смесь из k нормальных распределений с разными дисперсиями σ_i, $i = 1..k$.

Упражнение

Написать программу, реализующую эту версию EM; k и σ_i, $i = 1..k$, должно быть возможно задавать на вход.

Спасибо за внимание!

- Lecture notes, слайды и коды программ появятся на моей homepage:

  `http://logic.pdmi.ras.ru/~sergey/index.php?page=teaching`

- Присылайте любые замечания, коды программ на других языках, решения упражнений, новые численные примеры и прочее по адресам:
  `sergey@logic.pdmi.ras.ru`, `smartnik@inbox.ru`