Второй семинар.
ММП, осень 2012–2013
11 сентября

Темы семинара:

- работа с вероятностью, формула Байеса, формула полной вероятности;
- вероятностная постановка задачи обучения по прецедентам, минимизация среднего риска, МЭР; оценки максимального правдоподобия.
- нормальное распределение, свойства, оценки максимального правдоподобия его параметров.

1 Первая половина

Задача 1. Пусть имеется две корзины: красная (r) и синяя (b). В красной корзине лежит 6 апельсинов (o) и 2 яблока (a), а в синей — 3 яблока и один апельсин. Пусть мы вытягиваем фрукт из этих корзин следующим образом: сначала случайным образом выбираем корзину, из которой будем тянуть, а затем вытягиваем из выбранной корзины случайно один фрукт. Будем обозначать корзину буквой B, а фрукт — буквой F. При этом пусть вероятность выбрать красную корзину на первом шаге равна 0,4: $P(B = r) = 0{,}4$. На втором шаге фрукты из выбранной корзины вытягиваются с равными вероятностями.
а) Вычислите все значения условного распределения $P(F|B)$, $F \in \{o, a\}$, $B \in \{r, b\}$.
б) Является ли функция $f(Y) = P(X|Y)$ при фиксированном X распределением?
в) Вычислите все значения $P(F)$, $F \in \{o, a\}$, $B \in \{r, b\}$.
г) Вычислите $P(B = r|F = o)$ и $P(B = r|F = a)$. Вспомните про б).

Вспоминаем определения и полезные свойства математического ожидания дисперсии и ковариации:
$\mathsf{E}\{X + Y\} = \mathsf{E}Y + \mathsf{E}X$*;*
$\mathsf{E}\{XY\} = \mathsf{E}X\mathsf{E}Y$*, если* X *и* Y *независимы;*
$\mathsf{D}\{X + Y\} = \mathsf{D}\,X + \mathsf{D}\,Y$*, если* X *и* Y *независимы; (***вывести***)*
$\mathsf{E}_{X,Y} f(X, Y) = \mathsf{E}_X \mathsf{E}_{Y|X} f(X, Y) = \mathsf{E}_Y \mathsf{E}_{X|Y} f(X, Y)$ *— conditioning (***обосновать***);*
$cov(X, Y) = \mathsf{E}\{(X - \mathsf{E}X)(Y - \mathsf{E}Y)\} = \mathsf{E}\{XY\} - \mathsf{E}X\mathsf{E}Y = 0$*, если* X *и* Y *независимы.*

Напоминание вероятностной постановки задачи обучения по прецедентам. Коротко обсуждаем суть априорных вероятностей классов и условных вероятностей

$P(X|Y)$ и $P(Y|X)$. Еще одно обоснование МЭР — желание решить задачу минимизации риска:

$$\mathsf{E}_{X,Y}\{L\big(a(X),Y\big)\} \to \min_a,$$

где L — функция потерь. Поскольку мы не знаем $P(X,Y)$, можно воспользоваться законом больших чисел и перейти к решению минимизации эмпирического риска:

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(Y_i - a(X_i)\big) \to \min\nolimits_a .$$

Задача 2. Пусть в задаче регрессии ($Y = \mathbb{R}$) мы используем квадратичную функцию потерь $L(a,b) = (a-b)^2$. Доказать, что решением описанной задачи минимизации риска будет *функция регрессии* $a(X) = \mathsf{E}\{Y|X\}$.

2 Вторая половина

Напоминаю, что такое функция правдоподобия (данных при условии параметров модели), метод максимального правдоподобия и как выглядит одномерное нормальное распределение:

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}.$$

Задача 3. Найдите оценку максимального правдоподобия для математического ожидания одномерного нормального распределения. Смещена?

Напоминаю вид многомерного нормального распределения, смысл и свойства матрицы ковариаций и вектора среднего.

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\Sigma) = \frac{1}{\sqrt{2\pi}^n\sqrt{\det\Sigma}}\exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big), \ \ \boldsymbol{x}\in\mathbb{R}^n$$

Напоминаю ряд свойств матриц и производной по матрице. Алгебраическое дополнение элемента $a_{i,j}$ квадратной матрицы $A \in \mathbb{R}^{n,n}$: $A_{i,j} = (-1)^{i+j}\det\bar{A}^{i,j}$, где $\bar{A}^{i,j}$ — матрица, полученная из A выбрасыванием i-й строки и j-го столбца. Тогда определитель можно разложить по i-й строке: $\det A = \sum_{j=1}^n a_{i,j}A_{i,j}$. Присоединенная матрица — транспонированная матрица алгебраических дополнений: $\{C^*(A)\}_{i,j} = A_{j,i}$. Справедлива следующая формула: $A^{-1} = \frac{1}{\det A}C^*(A)$, а значит $(A^{-1})_{i,j} = \frac{A_{j,i}}{\det A}$.

Произвольная A	*Симметричная A*
(1) $\frac{\partial}{\partial u}u^\top Au = A^\top u + Au$	*(4)* $\frac{\partial}{\partial u}u^\top Au = 2Au$
(2) $\frac{\partial}{\partial A}\ln\det A = (A^{-1})^\top$	*(5)* $\frac{\partial}{\partial A}\ln\det A = 2(A^{-1}) - \operatorname{diag}A^{-1}$
(3) $\frac{\partial}{\partial A}u^\top Au = uu^\top$	*(6)* $\frac{\partial}{\partial A}u^\top Au = 2uu^\top - \operatorname{diag}uu^\top$.

Докажите (1), (4) и (6) дома самостоятельно.

Докажем (2). Поскольку $A_{i,j}$ *не зависит от элементов соответствующего столбика и строки, то с учетом сказанного, получаем* $\frac{\partial \det A}{\partial a_{i,j}} = A_{i,j} = (A^{-1})_{j,i} \det A$, *или* $\frac{\partial \det A}{\partial A} = \det A (A^{-1})^\top$. *А значит* $\frac{\partial \ln \det A}{\partial A} = \frac{1}{\det A} \det A (A^{-1})^\top = (A^{-1})^\top$.

Докажем (5), что немного сложнее. В данном случае у нас $n(n-1)/2$ *аргументов, что придется брать в расчет. Представим, что у нас есть матрица* B, *элементы* $b_{i,j}$ *и* $b_{j,i}$ *которой зависят от* $a_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, n$. *Тогда*

$$\begin{cases} \frac{\partial \det B}{\partial a_{i,j}} = \frac{\partial \det B}{\partial b_{i,j}} \frac{\partial b_{i,j}}{\partial a_{i,j}} + \frac{\partial \det B}{\partial b_{j,i}} \frac{\partial b_{j,i}}{\partial a_{i,j}} = B_{i,j} \frac{\partial b_{i,j}}{\partial a_{i,j}} + B_{j,i} \frac{\partial b_{j,i}}{\partial a_{i,j}}, & i \neq j \\ \frac{\partial \det B}{\partial a_{i,i}} = B_{i,i} \frac{\partial b_{i,i}}{\partial a_{i,i}}, & i = j. \end{cases}$$

Положив $b_{i,j} = b_{j,i} = a_{i,j} = a_{j,i}$, *с учетом* $A_{i,j} = A_{j,i}$ *получим окончательно*

$$\begin{cases} \frac{\partial \det A}{\partial a_{i,j}} = 2A_{i,j}, & i \neq j \\ \frac{\partial \det A}{\partial a_{i,i}} = A_{i,i}, & i = j. \end{cases} \text{ и как следствие:}$$

$$\frac{\partial}{\partial A} \ln \det A = 2A^{-1} - \operatorname{diag} A^{-1}.$$

Докажем (3). $u^\top A u = \sum_i \sum_j u_i a_{i,j} u_j$. *Производная по* $a_{i,j}$ *равна* $u_i u_j$. *Чтд.*

Задача 4. Получите оценки максимального правдоподобия для параметров n-мерного нормального распределения.

3 Решения

Задача 1:

а) В данном случае (с конечным числом значений сл. величин) можно пользоваться правилом $P(X = x|Y = y)$ — доля исходов с $X = x$ среди тех, у которых $Y = y$. (В более общем случае так не получится.)

$$P(F = o|B = b) = \frac{1}{4}; \tag{1}$$

$$P(F = a|B = b) = 1 - \frac{1}{4} = \frac{3}{4}; \tag{2}$$

$$P(F = o|B = r) = \frac{6}{8} = \frac{3}{4}; \tag{3}$$

$$P(F = a|B = r) = 1 - \frac{3}{4} = \frac{1}{4}. \tag{4}$$

б) Нет, эта функция не нормирована. Это мы увидим в пункте г).

в) По *формуле полной вероятности*

$$P(F = a) = P(F = a|B = b)P(B = b) + P(F = a|B = r)P(B = r) = \\ = \frac{3}{4} \times \frac{6}{10} + \frac{1}{4} \times \frac{4}{10} = \frac{11}{20}.$$

Учитывая условие нормировки $P(F=a)+P(F=o)=1$, получаем $P(F=o)=\frac{9}{20}$.
г) Используя *формулу Байеса* получаем:

$$P(B=r|F=o)=\frac{P(F=o|B=r)P(B=r)}{P(F=o)}=\frac{3}{4}\times\frac{4}{10}\times\frac{20}{9}=\frac{2}{3};$$

$$P(B=r|F=a)=\frac{P(F=a|B=r)P(B=r)}{P(F=a)}=\frac{1}{4}\times\frac{4}{10}\times\frac{20}{11}=\frac{2}{11}.$$

Задача 2: применим хитрый трюк:

$$\mathsf{E}\big(a(X)-Y\big)^2=\mathsf{E}\big(a(X)-\mathsf{E}\{Y|X\}+\mathsf{E}\{Y|X\}-Y\big)^2=$$
$$=\mathsf{E}\big(a(X)-\mathsf{E}\{Y|X\}\big)^2+\mathsf{E}\big(\mathsf{E}\{Y|X\}-Y\big)^2+2\mathsf{E}\Big\{\big(a(X)-\mathsf{E}\{Y|X\}\big)\big(\mathsf{E}\{Y|X\}-Y\big)\Big\}.$$

С помощью conditioning легко показать, что последний член равен нулю, поскольку $a(X)-\mathsf{E}\{Y|X\}$ не зависит от Y и $\mathsf{E}_{Y|X}\big(\mathsf{E}\{Y|X\}-Y\big)=0$.

Задача 3: Функция правдоподобия:

$$L(X^\ell|\,\mu,\sigma^2)=\prod_{i=1}^{\ell}\mathcal{N}(X_i|\mu,\sigma^2)\to\max_{\mu,\sigma^2};$$

$$\ln L(X^\ell|\,\mu,\sigma^2)=\ln\prod_{i=1}^{\ell}\mathcal{N}(X_i|\mu,\sigma^2)=\sum_{i=1}^{\ell}\ln\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}\to\max_{\mu,\sigma^2};$$

$$\begin{cases}\frac{\partial\ln L(X^\ell|\,\mu,\sigma^2)}{\partial\mu}=0;\\ \frac{\partial\ln L(X^\ell|\,\mu,\sigma^2)}{\partial\sigma^2}=0.\end{cases}\qquad\frac{\partial\ln L(X^\ell|\,\mu,\sigma^2)}{\partial\mu}=-\sum_{i=1}^{\ell}\frac{(X_i-\mu)}{\sigma^2}\dots$$

Задача 4: Пусть у нас есть ℓ наблюдений n-мерной нормальной случайной величины $\boldsymbol{X}=\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_\ell\}$. Для начала найдем μ_{MP}.

$$\ln\mathcal{N}(\boldsymbol{X}|\boldsymbol{\mu},\Sigma)=\frac{1}{2}\ln\det\Sigma^{-1}-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})+\text{const}(\boldsymbol{\mu},\Sigma);$$

$$\frac{\partial}{\partial\boldsymbol{\mu}}\ln L(\boldsymbol{X}|\boldsymbol{\mu},\Sigma)=\frac{\partial}{\partial\boldsymbol{\mu}}\sum_{i=1}^{\ell}-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})=\sum_{i=1}^{\ell}\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})=\Sigma^{-1}\sum_{i=1}^{\ell}(\boldsymbol{x}_i-\boldsymbol{\mu})=0$$

поскольку матрица Σ^{-1} невырождена, получаем $\boldsymbol{\mu}_{\text{MP}}=\frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i$.

Теперь найдем Σ_{MP}:

$$\frac{\partial}{\partial\Sigma^{-1}}\ln L(\boldsymbol{X}|\boldsymbol{\mu},\Sigma)=\ell\frac{1}{2}(2\Sigma-\text{diag}\Sigma)-\frac{1}{2}\sum_{i=1}^{\ell}\big(2(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top-\text{diag}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\big)=0$$

Обозначим $S = \ell\,\Sigma - \sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top$, тогда

$$\frac{\partial}{\partial \Sigma^{-1}} \ln L(\boldsymbol{X}|\boldsymbol{\mu}, \Sigma) = \frac{1}{2}(2S - \mathrm{diag} S)) = 0,$$

откуда получаем

$$S = \ell\,\Sigma - \sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top = 0$$

и окончательно

$$\Sigma_{\text{МР}} = \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top.$$