Tsinghua-Berkeley Shenzhen Institute
INFERENCE AND INFORMATION
Fall 2017

Homework 1

赵丰 November 21, 2017

- **Acknowledgments:** This coursework referes to wikipedia: `https://en.wikipedia.org`.
- **Collaborators:** I finish this template by myself.

I use `enumerate` to generate answers for each question:

1.1. (a) i.

$$\mathbb{E}[\mathsf{x}|\mathsf{y}] = \sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{y}) \tag{1}$$

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{yz}]|\mathsf{y}] =& \mathbb{E}((\sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{yz}))|\mathsf{y}) \\
=& \sum_j (\sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{y}, \mathsf{z} = z_j)) \Pr(\mathsf{z} = z_j) \\
=& \sum_i x_i (\sum_j \Pr(\mathsf{x} = x_i|\mathsf{y}, \mathsf{z} = z_j)) \Pr(\mathsf{z} = z_j) \\
=& \sum_i x_i (\sum_j \Pr(\mathsf{x} = x_i, \mathsf{z} = z_j|\mathsf{y})) \\
=& \sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{y})
\end{aligned}$$

所以

$$\mathbb{E}[\mathsf{x}|\mathsf{y}] = \mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{yz}]|\mathsf{y}] \tag{2}$$

ii.

$$g(\mathsf{y})\,\mathbb{E}[\mathsf{x}|\mathsf{y}] = \sum_i g(\mathsf{y}) x_i \Pr(\mathsf{x} = x_i|\mathsf{y}) = \mathbb{E}[x g(\mathsf{y})|\mathsf{y}] \tag{3}$$

iii.

$$\begin{aligned}
\mathbb{E}[\mathsf{x}\,\mathbb{E}[\mathsf{x}|\mathsf{y}]] =& \mathbb{E}\left[\mathsf{x} \sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{y})\right] \\
=& \sum_{j,k} (x_j \sum_i x_i \Pr(\mathsf{x} = x_i|\mathsf{y} = y_k)) \Pr(\mathsf{x} = x_j, \mathsf{y} = y_k) \\
=& \sum_{i,j,k} x_i x_j \Pr(\mathsf{x} = x_i|\mathsf{y} = y_k) \Pr(\mathsf{x} = x_j, \mathsf{y} = y_k)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[(\mathbb{E}[\mathsf{x}|\mathsf{y}])^2] &= \mathbb{E}\left[(\sum_i x_i \Pr(\mathsf{x}=x_i|\mathsf{y}))^2\right] \\
&= \sum_k (\sum_i x_i \Pr(\mathsf{x}=x_i|\mathsf{y}=y_k))^2 \Pr(\mathsf{y}=y_k) \\
&= \sum_{i,j,k} x_i x_j \Pr(\mathsf{x}=x_i|\mathsf{y}=y_k) \Pr(\mathsf{x}=x_j|\mathsf{y}=y_k) \Pr(\mathsf{y}=y_k) \\
&= \sum_{i,j,k} x_i x_j \Pr(\mathsf{x}=x_i|\mathsf{y}=y_k) \Pr(\mathsf{x}=x_j, \mathsf{y}=y_k)
\end{aligned}
$$

所以

$$
\mathbb{E}[\mathsf{x}\,\mathbb{E}[\mathsf{x}|\mathsf{y}]] = \mathbb{E}[(\mathbb{E}[\mathsf{x}|\mathsf{y}])^2] \tag{4}
$$

iv. 对离散型随机变量证明全方差公式 (Law of Total Variance) 用到全期望公式 (Law of Total Expectation)：

$$
\mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{y}]] = \mathbb{E}[\mathsf{x}] \tag{5}
$$

$$
\mathrm{Var}(\mathsf{x}) = \sum_i (x_i - \mathbb{E}(\mathsf{x}))^2 \Pr(\mathsf{x}=x_i) \tag{6}
$$

$$
\begin{aligned}
\mathrm{Var}(\mathsf{x}) &= \mathbb{E}[\mathsf{x}^2] - (\mathbb{E}[\mathsf{x}])^2 \\
&= \mathbb{E}[\mathbb{E}[\mathsf{x}^2|\mathsf{y}]] - (\mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{y}]])^2 \\
&= \mathbb{E}[\mathrm{Var}[\mathsf{x}|\mathsf{y}] + (\mathbb{E}[\mathsf{x}|\mathsf{y}])^2] - (\mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{y}]])^2 \\
&= \mathbb{E}[\mathrm{Var}[\mathsf{x}|\mathsf{y}]] + \mathrm{Var}[\mathbb{E}[\mathsf{x}|\mathsf{y}]]
\end{aligned}
$$

(b) i.

$$
\mathbb{E}[(\mathsf{y}-\alpha)^2] = \alpha^2 - 2\,\mathbb{E}[\mathsf{y}]\alpha + \mathbb{E}[\mathsf{y}]^2 \tag{7}
$$

由二次函数极值的性质，当 $\alpha = \mathbb{E}[\mathsf{y}]$ 时 $\mathbb{E}[(\mathsf{y}-\alpha)^2]$ 最小，此时 $\mathrm{MSE}(\hat{\mathsf{y}} = \mathbb{E}[\mathsf{y}]) = \mathrm{Var}(\mathsf{y})$

ii.

$$
\mathbb{E}[(\mathsf{y}-\hat{\mathsf{y}})^2] = \sum_i \left(\sum_j (y_j - \hat{\mathsf{y}}(x_i))^2 \Pr(\mathsf{y}=y_j|\mathsf{x}=x_i) \right) \Pr(\mathsf{x}=x_i) \tag{8}
$$

故只需对每一个 $\Pr(\mathsf{x}=x_i)$，极小化

$$
\left(\sum_j (y_j - f(x_i))^2 \Pr(\mathsf{y}=y_j|\mathsf{x}=x_i) \right) \tag{9}
$$

目标函数可化为如下形式:

$$
\mathcal{F}(f) = f(x_i)^2 - 2\,\mathbb{E}[\mathsf{y}|\mathsf{x}=x_i] f(x_i) + c, f:\mathcal{X} \to \mathbb{R} \tag{10}
$$

由第一问结论:

$$
\arg\min_{f:\mathcal{X}\to\mathbb{R}} \mathcal{F}(f) = \mathbb{E}[\mathsf{y}|\mathsf{x}=x_i] \tag{11}
$$

从而有

$$\arg\min_{f:\mathcal{X}\to\mathbb{R}} \mathbb{E}[(\mathsf{y}-f(\mathsf{x}))^2]=\mathbb{E}[\mathsf{y}|\mathsf{x}] \tag{12}$$

由(8)式得估计的 MSE 为

$$\mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}])=\sum_i \mathrm{Var}(\mathsf{y}|\mathsf{x}=x_i)\Pr(\mathsf{x}=x_i)=\mathbb{E}[\mathrm{Var}(\mathsf{y}|\mathsf{x})] \tag{13}$$

iii. 由全方差公式

$$\mathbb{E}[\mathrm{Var}(\mathsf{y}|\mathsf{x})]=\mathrm{Var}(\mathsf{y})-\mathrm{Var}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \tag{14}$$

若 x 与 y 相互独立，$\mathbb{E}[\mathsf{y}|\mathsf{x}]=\mathbb{E}[\mathsf{y}]$ 为常数，因此 $\mathrm{Var}(\mathbb{E}[\mathsf{y}|\mathsf{x}])=0$, 所以 $\mathbb{E}[\mathrm{Var}(\mathsf{y}|\mathsf{x})]=\mathrm{Var}(\mathsf{y})$, 即

$$\mathrm{MSE}(\mathbb{E}[\mathsf{y}])=\mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \tag{15}$$

若上式成立，可知 $\mathbb{E}[\mathsf{y}|\mathsf{x}]=\mathbb{E}[\mathsf{y}]$ 为常数，即

$$\forall i, \sum_j y_j \Pr(\mathsf{y}=y_j|\mathsf{x}=x_i)=\sum_j y_j \Pr(\mathsf{y}=y_j) \tag{16}$$

上式两边同时乘以 $\mathbb{E}[f(\mathsf{x})]$ 可得

$$\sum_{i,j} f(x_i)y_j \Pr(\mathsf{x}=x_i,\mathsf{y}=y_j)=\left(\sum_i f(x_i)\Pr(\mathsf{x}=x_i)\right)\left(\sum_j y_j\Pr(\mathsf{y}=y_j)\right) \tag{17}$$

即 $\mathbb{E}[f(\mathsf{x})\mathsf{y}]=\mathbb{E}[f(\mathsf{x})]\,\mathbb{E}[\mathsf{y}]$ 所以 $\forall f, \rho(f(\mathsf{x}),\mathsf{y})=0$

1.2. (a)

$$\mathbb{E}[(\mathsf{y}-a\mathsf{x}-b)^2]==a^2\,\mathbb{E}[\mathsf{x}^2]+\mathbb{E}[\mathsf{y}^2]+b^2-2a\,\mathbb{E}[\mathsf{xy}]-2b\,\mathbb{E}[\mathsf{y}]+2ab\,\mathbb{E}[\mathsf{x}] \tag{18}$$

对 a, b 求偏导得：

$$\begin{cases}\mathbb{E}[\mathsf{x}^2]a+\mathbb{E}[\mathsf{x}]b= & \mathbb{E}[\mathsf{xy}]\\ \mathbb{E}[\mathsf{x}]a+b= & \mathbb{E}[\mathsf{y}]\end{cases} \tag{19}$$

从而解出，并考虑到 $\mathrm{Var}(\mathsf{x})=\mathrm{Var}(\mathsf{y})=\sqrt{\mathrm{Var}(\mathsf{x})\,\mathrm{Var}(\mathsf{y})}$:

$$a=\frac{\mathbb{E}[\mathsf{xy}]-\mathbb{E}[\mathsf{x}]\,\mathbb{E}[\mathsf{y}]}{\mathrm{Var}(\mathsf{x})}=\rho(\mathsf{x},\mathsf{y}) \tag{20}$$

(b)

$$\mathsf{x}\perp\mathsf{y}\iff\Pr(\mathsf{x}=x_i,\mathsf{y}=y_j)=\Pr(\mathsf{x}=x_i)\Pr(\mathsf{y}=y_j) \tag{21}$$

所以当 x 与 y 独立时

$$\begin{aligned}\mathbb{E}[f(\mathsf{x})g(\mathsf{y})]&=\sum_{i,j} f(x_i)g(\mathsf{y}_j)\Pr(\mathsf{x}=x_i,\mathsf{y}=y_j)\\&=\sum_{i,j} f(x_i)g(y_j)\Pr(\mathsf{x}=x_i)\Pr(\mathsf{y}=y_j)\\&=\left(\sum_i f(x_i)\Pr(\mathsf{x}=x_i)\right)\left(\sum_i g(y_j)\Pr(\mathsf{y}=x_j)\right)\\&=\mathbb{E}[f(\mathsf{x})]\,\mathbb{E}[g(\mathsf{y})]\end{aligned}$$

即 $\forall f, g, \rho(f(\mathsf{x}), g(\mathsf{y})) = 0$
另一方面，若 $\forall f, g, \rho(f(\mathsf{x}), g(\mathsf{y})) = 0$。取 $f = 1_{\mathsf{x}=x_i}, g = 1_{\mathsf{y}=y_j}$，则由 $\mathbb{E}[f(\mathsf{x})g(\mathsf{y})] = \mathbb{E}[f(\mathsf{x})]\,\mathbb{E}[g(\mathsf{y})]$ 可以推出

$$\Pr(\mathsf{x} = x_i, \mathsf{y} = y_j) = \Pr(\mathsf{x} = x_i)\Pr(\mathsf{y} = y_j) \tag{22}$$

所以 x 与 y 相互独立。

1.3. Thanks to 陆石, who gives me this template.