IML 第一次作业

作业 1.1

$$
\begin{aligned}
\frac{\partial \ln \det(A)}{\partial x} &= \frac{1}{\det(A)} \frac{\partial \det(A)}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \det(A)}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \left[\sum_{k=1}^{n} A_{kj} C_{kj}\right]}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial A_{kj} C_{kj}}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left(\frac{\partial A_{kj}}{\partial A_{ij}} \cdot C_{kj} + \frac{\partial C_{kj}}{\partial A_{ij}} A_{kj} \right) \cdot \frac{\partial A_{ij}}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(C_{ij}\right) \frac{\partial A_{ij}}{\partial x} \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left[adj(A)_{ji}\right] \frac{\partial A_{ij}}{\partial x} \right) \\
&= \frac{1}{\det(A)} \sum_{i=1}^{n} \left(\frac{\partial A}{\partial x} \cdot adj(A) \right)_{ii} \\
&= \frac{1}{\det(A)} \operatorname{tr} \left(\frac{\partial A}{\partial x} \cdot adj(A) \right) \\
&= \operatorname{tr} \left(\frac{\partial A}{\partial x} \cdot \frac{adj(A)}{\det(A)} \right) = \operatorname{tr} \left(\frac{\partial A}{\partial x} \cdot A^{-1} \right)
\end{aligned}
$$

作业 1.2

色泽：青绿、乌黑（共 2 种）

根蒂：蜷缩、硬挺、稍蜷（共 3 种）

敲声：浊响、清脆、沉闷（共 3 种）

单合取式构成的假设空间有 $3 * 4 * 4 + 1 = 49$ 种。若使用最多包含 k 个合取式的析合范式来表达假设空间，需要从 48 种非空假设中选择并去除冗余情况。下面用概率的方法来估计。

每次选一个合取式，其可能会兼并其他假设的概率为

$$p_1 = 1 - \frac{2 * 3 * 3 - 1}{3 * 4 * 4 - 1} = 17/47$$

可能兼并其他假设的合取式共有 $48 - 18 - 1 = 29$ 种，其中可能兼并 2 种假设的有 9 种，可能兼并 3 种假设的有 12 种，可能兼并 5 种假设的有 6 种，可能

兼并 6 种假设的有 2 种。

某合取式子兼并其他假设的期望是

$$E = \sum p(\text{兼并})N(\text{兼并假设数}) = 3.3$$

若不兼并，最多 k 个合取支构造的假设数为

$$N_1 = \sum_{i=1}^{k} C_{48}^{i}$$

若把每次选择看成是随机选择的，由于这样选择则兼并的假设总数是

$$N_2 = cN_1p_1E$$

其中 c 是随机选择与全部选择差异导致的系数，当 k 很大时 c 可以取 1/3，这里用指数函数来模拟这一变化。

估计如下：

$k = 1, N_2 = 48$

$k = 2, N_2 = 1123$

$k = 3, N_2 = 17246$

$k = 4, N_2 = 194423$

$k = 5, N_2 = 1717270$

$k = 6, N_2 = 12376152$

$k = 7, N_2 = 74830982$

$k = 8, N_2 = 387382919$

$k = 9, N_2 = 1743677791$

$k = 10, N_2 = 6907563381$

$k = 11, N_2 = 24319937067$

$k = 12, N_2 = 76714685703$

$k = 13, N_2 = 218280218709$

$k = 14, N_2 = 563492921612$

$k = 15, N_2 = 1326471675187$

$k = 16, N_2 = 2860217495798$

$k = 17, N_2 = 5672482962030$

$k = 18, N_2 = 10386892218357$

$k = 19, N_2 = 17625423696294$

$k = 20, N_2 = 27818228259221$

$k = 21, N_2 = 40991617896730$

$k = 22, N_2 = 56620929422966$

$k = 23, N_2 = 73633037038712$
$k = 24, N_2 = 90591671446655$
$k = 25, N_2 = 106017488157596$
$k = 26, N_2 = 118728553208919$
$k = 27, N_2 = 128075835963284$
$k = 28, N_2 = 133999547522100$
$k = 29, N_2 = 136914425552286$
$k = 30, N_2 = 137498275817139$
$k = 31, N_2 = 136477279037359$
$k = 32, N_2 = 134475260955399$
$k = 33, N_2 = 131947966117278$
$k = 34, N_2 = 129185515384135$
$k = 35, N_2 = 126350575714537$
$k = 36, N_2 = 123523692061307$
$k = 37, N_2 = 120739854896841$
$k = 38, N_2 = 118012024841509$
$k = 39, N_2 = 115343830530051$
$k = 40, N_2 = 112735442211470$
$k = 41, N_2 = 110185921158588$
$k = 42, N_2 = 107694034252687$
$k = 43, N_2 = 105258498150961$
$k = 44, N_2 = 102878041946684$
$k = 45, N_2 = 100551420493746$
$k = 46, N_2 = 98277416360327$
$k = 47, N_2 = 96054839593486$
$k = 48, N_2 = 93882527146425$
数字可能估计的不准，但是趋势是先增加再减少应该没错。

作业 1.3

设 $\boldsymbol{x}$ 是 n 维向量, $\boldsymbol{x}_1$ 是 m 维向量, 相应地 $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2], \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{mm} & \boldsymbol{\Sigma}_{m(n-m)} \\ \boldsymbol{\Sigma}_{(n-m)m} & \boldsymbol{\Sigma}_{(n-m)(n-m)} \end{bmatrix}$,
则 $P(\boldsymbol{x}_1) = \frac{1}{\sqrt{(2\pi)^m \det \boldsymbol{\Sigma}_{mm}}} e^{-\frac{1}{2}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{mm}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)}$,
$P(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$
$P(\boldsymbol{x}_2) = \frac{1}{\sqrt{(2\pi)^{n-m} \det \boldsymbol{\Sigma}_{(n-m)(n-m)}}} e^{-\frac{1}{2}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{(n-m)(n-m)}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)}$
$P(\boldsymbol{x}_1 \mid \boldsymbol{x}_2) = \frac{P(\boldsymbol{x})}{P(\boldsymbol{x}_2)} = \sqrt{\frac{(2\pi)^{-m} \det \boldsymbol{\Sigma}_{(n-m)(n-m)}}{\det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})+\frac{1}{2}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{(n-m)(n-m)}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)}$

作业 1.4

引理 (Hölder) 设 $p>1$，$\frac{1}{p}+\frac{1}{q}=1$ 。令 $a_1,\dots a_n$ 和 $b_1,\dots,b_n$ 是非负实数。那么

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}$$

证明 $f(\boldsymbol{x})=\|\boldsymbol{x}\|_p=\sqrt[p]{\sum_{i=1}^{r}|x_i|^p}$, 即证 $\forall t\in[0,1], \forall \boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}(f)$, 有

$$f(t\boldsymbol{x}+(1-t)\boldsymbol{y})\leq tf(\boldsymbol{x})+(1-t)f(\boldsymbol{y})$$

设 $\frac{1}{p}+\frac{1}{q}=1(p,q>1)$, 则

LHS $=\sqrt[p]{\sum_{i=1}^{r}|tx_i+(1-t)y_i|^p}=\sqrt[p]{\sum_{i=1}^{r}|tx_i+(1-t)y_i|\,|tx_i+(1-t)y_i|^{p-1}}$

$\leq \sqrt[p]{\sum_{i=1}^{r}\left[|tx_i|+|(1-t)y_i|\right]|tx_i+(1-t)y_i|^{p-1}}$

$=\sqrt[p]{\sum_{i=1}^{r}|tx_i|\,|tx_i+(1-t)y_i|^{p-1}+\sum_{i=1}^{r}|(1-t)y_i|\,|tx_i+(1-t)y_i|^{p-1}}$

$\leq \sqrt[p]{\left(\sum_{i=1}^{r}|tx_i|^p\right)^{\frac{1}{p}}k+\left(\sum_{i=1}^{r}|(1-t)y_i|^p\right)^{\frac{1}{p}}k}$

$=\sqrt[p]{\left[\left(\sum_{i=1}^{r}|tx_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{r}|(1-t)y_i|^p\right)^{\frac{1}{p}}\right]\left(\sum_{i=1}^{r}|tx_i+(1-t)y_i|^p\right)^{\frac{1}{q}}}$

$=\sqrt[p]{\left(\|t\boldsymbol{x}\|_p+\|(1-t)\boldsymbol{y}\|_p\right)\left(\sum_{i=1}^{r}|tx_i+(1-t)y_i|^p\right)^{\frac{1}{q}}}$

即 $\sqrt[p]{\sum_{i=1}^{r}|tx_i+(1-t)y_i|^p}\leq\sqrt[p]{\left(\|t\boldsymbol{x}\|_p+\|(1-t)\boldsymbol{y}\|_p\right)\left(\sum_{i=1}^{r}|tx_i+(1-t)y_i|^p\right)^{\frac{1}{q}}}$。

其中 $k=\left(\sum_{i=1}^{r}|tx_i+(1-t)y_i|^{(p-1)q}\right)^{\frac{1}{q}}$。

注意到左边根号下的项和右边根号下乘积第二项是相同的, 除过去再两边求 p 次幂即得证。

作业 1.5

必要性:

$f(tx+(1-t)y)\leq tf(x)+(1-t)f(y)$

$\Rightarrow f((1-t)(y-x)+x)\leq f(x)+(1-t)(f(y)-f(x))$

$\Rightarrow f((1-t)(y-x)+x)-f(x)\leq (1-t)(f(y)-f(x))$

当 $t=1{,}0$ 时, 此式恒成立, 对 $t\in(0{,}1)$, 不妨设 $x<y$, 有

$$\frac{f((1-t)(y-x)+x)-f(x)}{(1-t)(y-x)}\leq\frac{f(y)-f(x)}{y-x}$$

两边同时取 $t\to 1^-$, 即

$$\nabla f(x)^\top\leq\frac{f(y)-f(x)}{y-x}$$

整理即得一阶条件。

充分性：

令 $z = tx + (1-t)y$，则

$$f(y) \geq f(z) + \nabla f(z)^\top (y-z)$$

$$f(x) \geq f(z) + \nabla f(z)^\top (x-z)$$

假定 $x < y$，则有大小关系 $x < z < y$，考虑

$$f(y) \geq f(z) + \nabla f(z)^\top (y-z)$$

$$f(x) \geq f(z) + \nabla f(z)^\top (x-z)$$

分离梯度项

$$\frac{f(y)-f(z)}{y-z} \geq \nabla f(z)^\top$$

$$\frac{f(x)-f(z)}{x-z} \leq \nabla f(z)^\top$$

得

$$\frac{f(x)-f(z)}{x-z} \leq \frac{f(y)-f(z)}{y-z}$$

则

$$[f(x)-f(z)](y-z]) \geq [f(y)-f(z)](x-z)$$

则

$$yf(x) - zf(x) - xf(y) + zf(y) \geq yf(z) - xf(z)$$

则

$$t(y-x)f(x) + (1-t)(y-x)f(y) \geq (y-x)f(z)$$

两边同除 $y-x$（>0）即证。

习题 2.2

由于数据集中正反例各一半，因而在 10 折交叉验证中分层采样得到的每个子集数据的分布都是近似正反例各一半，根据学习模型的假定，最终的错误率大约是 $\frac{1}{2}$。

用留一法，假设每次取的测试样本是 A，则训练集中 $\bar{A}$ 较多，模型每次都会猜测是 $\bar{A}$，导致每次都猜错，最终的错误率大约是 1。

习题 2.4

真正例率：预测为正例且真实情况为正例的，占所有真实情况中正例的比率。

$$TPR = \frac{TP}{TP + FN}$$

假正例率：预测为正例但真实情况为反例的，占所有真实情况中反例的比率。

$$FPR = \frac{FP}{TN + FP}$$

查准率：预测结果为正例中实际正例的占比。

$$P = \frac{TP}{TP + FP}$$

查全率：真实情况为正例中预测正例的占比。

$$R = \frac{TP}{TP + FN}$$

可见 $TPR = R$，其他的比率之间没有直接的联系。

习题 2.5

曲线下方面积

$$AUC = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i+1} - x_i)(y_i + y_{i+1})$$

是用梯形的面积公式来计算的。而

$$l_{rank} = \frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}[\mathbb{I}(f(x^+) < f(x^-)) + \frac{1}{2}\mathbb{I}(f(x^+) = f(x^-))]$$

可以变形为

$$l_{rank} = \sum_{x^+\in D^+} S$$

其中

$$S = \frac{1}{2}\cdot\frac{1}{m^+}\left[\frac{2}{m^-}\sum_{x^-\in D^-}\mathbb{I}(f(x^+) < f(x^-)) + \frac{1}{m^-}\sum_{x^-\in D^-}\mathbb{I}(f(x^+) = f(x^-))\right]$$

是新增真正例导致的新增点与 y 轴围成的梯形面积。

梯形面积的高是 $\frac{1}{m^+}$，较短的底是 $\frac{1}{m^-}$ 乘以预测值大于 $f(x^+)$ 的假正例个数，较长的底是 $\frac{1}{m^-}$ 乘以预测值大于等于 $f(x^+)$ 的假正例个数。

$\sum_{x^+\in D^+}$ 在累计所有真正例导致的曲线变动。区域总面积为 1，则 $AUC = 1 - l_{rank}$。

习题 2.9

假设实验中从总体中随机取样得到的 n 个观察值被划分为 k 个互斥的分类，这样每个分类都有一个对应的实际观察次数 $x_i(i=1,2,\dots,k)$ 。对实验中各个观察值落入第 i 个分类的概率 p_i 的分布提出零假设，得所有第 i 分类的理论期望次数 $m_i = np_i$ 以及条件

$$\sum_{i=1}^{k} p_i = 1 \quad \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} x_i = n。$$

在上述零假设成立以及 n 趋向 ∞ 的时候，以下统计量的极限分布趋向 χ^2 分布。

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i} = \sum_{i=1}^{k} \frac{x_i^2}{m_i} - n$$

根据样本数据计算上述统计量，取显著水平为 α。将计算结果与 χ^2 分布的上 $1-\alpha$ 分点比较，然后选择是否推翻零假设。