第一周作业

T_1 求 $\frac{\partial \ln \det(A)}{\partial x}$

先求 $\frac{\partial \det(A)}{\partial A_{ij}} = \frac{\partial \sum_k A_{ik}\, adj^T(A)_{ik}}{\partial A_{ij}} = \sum_k \frac{\partial A_{ik}\, adj^T(A)_{ik}}{\partial A_{ij}}$

$$= \sum_k \frac{\partial A_{ik}}{\partial A_{ij}} adj^T(A)_{ik} + \sum_k A_{ik} \frac{\partial adj^T(A)_{ik}}{\partial A_{ij}}$$

$$= \sum_k \frac{\partial A_{ik}}{\partial A_{ij}} adj^T(A)_{ik} = adj^T(A)_{ij}.$$

$$\frac{\partial \ln \det(A)}{\partial x} = \frac{1}{\det(A)} \sum_{ij} \frac{\partial \det(A)}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial x}$$

$$= \frac{1}{\det(A)} \sum_{ij} adj(A)^T_{ij} \frac{\partial A_{ij}}{\partial x}$$

$$= \frac{1}{\det(A)} \sum_{ij} adj(A)_{ji} \frac{\partial A_{ij}}{\partial x}$$

$$= \frac{1}{\det(A)} \sum_{j=1}^{n} \left(A^* \frac{\partial A}{\partial x}\right)_{jj} = tr\left(A^{-1} \frac{\partial A}{\partial x}\right)$$

T_2. 色泽 — 青绿 / 乌黑 / * "*"表示这一属性取任意值都合适

根蒂 — 蜷缩 / 稍蜷 / 硬挺 / * 假设空间大小为 $3\times4\times4+1=49$

敲声 — 浊响 / 清脆 / 沉闷 / * （此处的1指空集）

不考虑空集：可能的假设一共 $\sum_{n=0}^{k} C_{48}^{n}$ （即从48种假设中任取n（$0\le n\le 49$）

k为最多允许的析取式数 种合取得到的所有情况）

2023.09.27

T3. $x=[x_1,x_2] \sim N(\mu,\Sigma)$ 求 $p(x_1)$，$p(x_1|x_2)$

$$N(x;\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \quad \rho=\frac{cov(x_1,x_2)}{\sigma_1\sigma_2} \quad \Sigma^{-1}=\frac{1}{(1-\rho^2)\sigma_1^2\sigma_2^2}\begin{pmatrix}\sigma_1^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}$$

$$=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} \qquad |\Sigma|=(1-\rho^2)\sigma_1\sigma_2$$

$$f(x_1,x_2)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^2|\Sigma|}}=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

其中 $(x-\mu)^T\Sigma^{-1}(x-\mu)=(x_1-\mu_1,x_2-\mu_2)\frac{1}{(1-\rho^2)\sigma_1^2\sigma_2^2}\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}(x_1-\mu_1,x_2-\mu_2)^T$

$$=\frac{1}{(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

$$p(x_1)=\int_{-\infty}^{+\infty}p(x_1,x_2)dx_2=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

令 $u=\frac{x_1-\mu_1}{\sigma_1}$，$v=\frac{x_2-\mu_2}{\sigma_2}$，得

① $$p(x_1)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}[u^2-2\rho uv+v^2]\right\}\sigma_1\sigma_2$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}}dv=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \quad (-\infty<x_1<\infty)$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{\Sigma_{11}}}e^{-\frac{(x_1-\mu_1)^2}{2\Sigma_{11}}} \quad (-\infty<x_1<\infty)$$

② 令 $z=x_1+Ax_2$ 其中 $A=-\Sigma_{12}\Sigma_{22}^{-1}$

$cov(z,x_2)=cov(x_1,x_2)+cov(Ax_2,x_2)=\Sigma_{12}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}=0$ 得z与x_2相互独立

$E(x_1|x_2)=E(z-Ax_2|x_2)=E(z|x_2)-E(Ax_2|x_2)=E(z)-Ax_2=\mu_1+A(\mu_2-x_2)=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$

$var(x_1|x_2)=var(z-Ax_1|x_2)=var(z|x_2)+var(Ax_2|x_2)-Acov(z,-x_2)-cov(z,-x_2)A'$

$=var(z|x_2)=var(z)=var(x_1+Ax_2)$

$=var(x_1)+Avar(x_2)A'+Acov(x_1,x_2)+cov(x_2,x_1)A'$

$=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

又∵多元正态分布的条件分布仍是多元正态的 ∴ $p(x_1|x_2)=N\left(x_1 \middle| \mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

T4. 证明

已

T5. 证

等价于

充

T4. 证明 $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$

已知. $\|x+y\|_p \leq \|x\|_p + \|y\|_p$ (由范数的性质知)

令 $f(x) = \|x\|_p$.

则 $t f(x) + (1-t) f(y) = t\|x\|_p + (1-t)\|y\|_p$

$$
\begin{aligned}
f(tx+(1-t)y) &= \|tx+(1-t)y\|_p \\
&\leq \|tx\|_p + \|(1-t)y\|_p \\
&= t\|x\|_p + (1-t)\|y\|_p \\
&= t f(x) + (1-t) f(y)
\end{aligned}
$$

则有: $\|x\|_p$ 是凸函数.

T5. 证明: 对 $\forall t \in [0,1]$

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

等价于 $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

充分性: 令 $z = tx+(1-t)y$.

已知 $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

则有: $\begin{cases} f(y) \geq f(z) + \nabla f(z)^T (z-y) \\ f(x) \geq f(z) + \nabla f(z)^T (z-x) \end{cases}$

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

$(1-t) f(y) \geq (1-t) f(z) + (1-t) \nabla f(z)^T (z-y)$

$$
\begin{aligned}
t f(x) + (1-t) f(y) &\geq f(z) + \nabla f(z)^T (t(1-t)(x-y) + t(1-t)(y-x)) \\
&= f(z)
\end{aligned}
$$

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

必要性: 已知 $f(tx+(1-t)y) \leq t f(x) + (1-t) f(y)$

$tx+(1-t)y = x + t(y-x)$

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

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

$f(y) \geq f(x) + \dfrac{f(x+t(y-x)) - f(x)}{t}$

当 $t \to 0$ 时, 即有 $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

综上, 证得 凸函数的 0阶和 1阶条件的等价性.

HW2

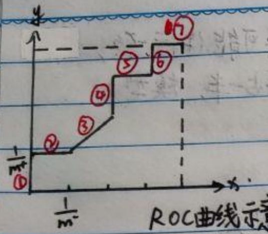
2.2 10折交叉验证：交叉验证中每个子集中数据分布要尽可能保持一致，则可认为10折交叉验证中10次训练中正反例各占一半，模型对新样本预测时随机猜测，错误率为50%。

留一法：取99个样本作训练集，1个样本作验证集
当这1个样本为正（反）例时，训练集中反（正）例更多，会作出（反）正的预测，错误率100%。

2.4 $TPR = \frac{TP}{TP+FN}$ $FPR = \frac{FP}{TN+FP}$ $P = \frac{TP}{TP+FP}$ $R = \frac{TP}{TP+FN}$

关系：$FPR + P = 1$（假正例率＋查准率＝1）
$TPR = R$.（真正例率＝查准率）

2.5 由AUC的定义知，AUC＝ROC曲线下的各部分面积求和.
又已知 ℓ_{rank} 定义为 $\frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}\left(\mathbb{I}(f(x^+)<f(x^-))+\frac{1}{2}\mathbb{I}(f(x^+)=f(x^-))\right)$
要证 $AUC = 1-\ell_{rank}$，即证.ROC曲线上的各部分面积和
等于 $\frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}\left(\mathbb{I}(f(x^+)<f(x^-))+\frac{1}{2}\mathbb{I}(f(x^+)=f(x^-))\right)$



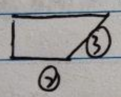
ROC曲线示意

如图是一个ROC曲线示意图，以 $\frac{1}{m^-}$ 为x轴的步长，以 $\frac{1}{m^+}$ 为y轴的步长。图中共有7段线段。其中①、④、⑥表示在分类阈值变动的过程中只新增了真正例，②、⑤表示在分类阈值变动的过程中只新增了假正例，③表示既新增了真正例也新增了反正例。

设这三类线段分别为A、B、C类

A、C类线段与y轴之间围成的面积可以用梯形面积的计算公式。

以③为例



↗即步长×预测值大于 $f(x^+)$ 的假正例的个数

上底：$\frac{1}{m^-}\sum_{x^-\in D^-} \mathbb{I}(f(x^+) < f(x^-))$
(短)

(长)
下底：预测值大于等于 $f(x^+)$ 的假正例的个数 × 步长

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

梯形的高：$\frac{1}{m^+}$

对①、④、⑥可以使用同样的梯形求解公式。

最后只需对所有这样的梯形求和：

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

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

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

$= \ell_{rank}$.

故有 ℓ_{rank} = ROC曲线与y轴围成的面积　由定义有 $AUC = 1 - \ell_{rank}$

2.9 χ^2检验过程：(1) 建立无关性假设，通过数据构建四格表.

(2) 根据假设生成新的理论四格表.

(3) 计算χ^2的值.

(4) 根据χ^2值查询卡方分布的临界值表，得出卡方检验的结果.

补充：如何按照比例对给定数据集作随机划分. 数据编号

① 获取数据集的数据总条数，并随机打乱，得到一个无序的数组

② 顺序划分数据集，并根据划分后每组数据中对应的数据的编号，在原始数据集中找到对应数据，实现分组。