HW1 & HW2 Reference

HW1

1. Calculate $\dfrac{\partial \ln\det(\mathbf{A})}{\partial x}$

Solve: 在我们的课程中，我们已经证明了

$$\frac{\partial \ln\det(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{A}^{-T}$$

右边的矩阵代表当 A_{ij} 变动时，$\ln\det(\mathbf{A})$ 将会变动 $(\mathbf{A}^{-T})_{ij}\cdot d\mathbf{A}_{ij}$.

我们又知道

$$\frac{\partial \mathbf{A}}{\partial x}$$

也是一个 $i\times j$ 的矩阵，当 x 变动时，$\mathbf{A}_{ij}$ 将变动 $(\dfrac{\partial \mathbf{A}}{\partial x})_{ij}\cdot dx$

综上，$\dfrac{\partial \ln\det(\mathbf{A})}{\partial x}$ 为以下矩阵的元素和

$$\mathbf{A}^{-T} * \frac{\partial \mathbf{A}}{\partial x}$$

其中 $*$ 代表 Hadamard Product，即两个矩阵对应元素乘积

因此最终结果可以表述为（这里可以）

$$\frac{\partial \ln\det(\mathbf{A})}{\partial x} = tr(\mathbf{A}^{-1}\cdot\frac{\partial \mathbf{A}}{\partial x})$$

另外，永停姐姐提供了一个证法，利用等式 $\ln(\det(A)) = tr(\ln(A))$。易得

$$\frac{\partial \ln\det(\mathbf{A})}{\partial x} = \frac{\partial tr(\ln(\mathbf{A}))}{\partial x} = tr(\frac{\partial \ln(\mathbf{A})}{\partial x}) = tr(\mathbf{A}^{-1}\cdot\frac{\partial \mathbf{A}}{\partial x})$$

其中 $\ln\mathbf{A}$ 满足 $\exp(\ln\mathbf{A}) = \mathbf{A}, \exp(\mathbf{M}) = \mathbf{I}+\mathbf{M}+\frac{1}{2}\mathbf{M}^2\dots$

2. 书习题1.2

Solve: 略，言之有理即可

3. 已知随机变量 $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2] \sim \mathcal{N}(\mu, \Sigma)$,计算 $P(\mathbf{x}_1), P(\mathbf{x}_1|\mathbf{x}_2)$

Solve: 我们先假设 $\mathbf{x}\in\mathbb{R}^n, \mathbf{x}_1\in\mathbb{R}^{n_1}, \mathbf{x}_2\in\mathbb{R}^{n_2}, n_1+n_2=n, \Sigma\in\mathbb{R}^{n\times n}, \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}, \mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$

由于 Σ 正定，顺序主子式 $|\Sigma_{11}|>0$, 因此 Σ_{11} 可逆，我们对 Σ 进行分解

$$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix} = \begin{pmatrix} I_{n_1} & 0\\ \Sigma_{12}^T\Sigma_{11}^{-1} & I_{n_2}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}-\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix}\begin{pmatrix} I_{n_1} & \Sigma_{11}^{-1}\Sigma_{12}\\ 0 & I_{n_2}\end{pmatrix}$$

取逆得到

$$\Sigma^{-1}=\begin{pmatrix} I_{n_1} & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I_{n_2}\end{pmatrix}\begin{pmatrix}\Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22}-\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\end{pmatrix}\begin{pmatrix} I_{n_1} & 0 \\ -\Sigma_{12}^T\Sigma_{11}^{-1} & I_{n_2}\end{pmatrix}$$

记 $\Sigma_{22}-\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}=\Sigma_*$

写出概率密度函数

$$f(x)=\frac{1}{(\sqrt{2\pi})^n\det(\Sigma)}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

对于指数部分，我们可以得到

$$\begin{aligned}(x-\mu)^T\Sigma^{-1}(x-\mu)=&(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)\\ &+[(x_2-\mu_2)-\Sigma_{12}^T\Sigma_{11}^{-1}(x_1-\mu_1)]^T\Sigma_*^{-1}[(x_2-\mu_2)-\Sigma_{12}^T\Sigma_{11}^{-T}(x_1-\mu_1)]\end{aligned}$$

当我们求 $P(x_1)$ 时候，实际是对 $x_2\in\mathbb{R}^{n_2}$ 求积分，由于 x_1 不变，可以认为求了一个 $\mathcal{N}(\mu_2+\Sigma_{12}^T\Sigma_{11}^{-1}(x_1-\mu_2),\Sigma_*)$ 的积分，可以求得对含 x_2 的指数部分求积分结果为 $|\Sigma|/|\Sigma_{11}|$, 也可以分析得出

$$x_1\sim\mathcal{N}(\mu_1,\Sigma_{11}).$$

当计算条件分布的时候，我们利用公式

$$P(x_1|x_2)=\frac{P(x_1,x_2)}{P(x_2)}$$

再利用之前得到的结果(注意需要将 Σ^{-1} 对于 Σ_{22} 分解)，可以得到结果

$$\begin{aligned}P(x_1|x_2)=\sqrt{\frac{|\Sigma_{22}|}{(2\pi)^{n_2}|\Sigma|}}\exp(-\frac{1}{2}([(x_1-\mu_1)^T-(x_2-\mu_2)^T\Sigma_{22}^{-1}\Sigma_{12}^T](\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}\\ [(x_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)])\end{aligned}$$

4. 证明 $\|x\|_p$ 是凸函数

   一个说明，$p<1,p\neq 1$ 时，不能满足向量范数要求的三角不等式，因此不能算范数，我们只需考虑 $p>1$ 的情况.（$p=0,p=1$ 显然）

Proof. 对于 $\forall t\in(0,1),u,v\in\mathbb{R}^n$, 有

$$\|tu+(1-t)v\|_p\le\|tu\|_p+\|(1-t)v\|_p=t\|u\|_p+(1-t)\|v\|_p$$

其中用到了向量范数必须满足的三角不等式和正齐次性

5. 证明判定凸函数的0阶和1阶条件相互等价

Proof. 充分性：设满足 $\forall t\in[0,1],f(tx+(1-t)y)\le tf(x)+(1-t)f(y)$

令 $g(t)=tf(x)+(1-t)f(y)-f(tx+(1-t)y)$, 显然 $g(0)=0,g(t)\ge 0,t\in[0,1]$. 易得 $\lim_{t\to 1-}g'(t)\le 0$

而

$$g'(t) = f(x) - f(y) - \nabla f(tx + (1-t)y)^T(x-y)$$
$$\lim_{t \to 1-} g'(t) = f(x) - f(y) - \nabla f(x)^T(x-y) \leq 0$$
$$f(y) \geq f(x) + \nabla f(x)^T(y-x)$$

必要性：设满足 $f(y) \geq f(x) + \nabla f(x)^T(y-x)$. 易得 $\lim_{t\to 0+} g'(t) \geq 0, \lim_{t \to 1-} \leq 0$. 倘若我们可以证明 $g''(t) \leq 0$, 0阶条件得证.

$$g''(t) = -(x-y)^T\nabla^2 f(tx + (1-t)y)(x-y)$$

只需证明 ∇f 是正定的即可. 而对于任意向量 x, y

$$f(y) - f(x) - \nabla f(x)^T(y-x) = (y-x)^T\nabla^2 f(x)(y-x) + O(\|y-x\|^3) \geq 0$$

当 $\|y-x\| \to 0$ 时，可以发现 $(y-x)^T\nabla^2 f(x)(y-x) \geq 0$, 由于 x, y 任意取，$\nabla^2 f$ 正定，证毕

HW2

1. 习题2.2

Solve：

10折交叉验证要求我们保证每个子集尽可能保持数据分布的一致性，即正反例数量相同，所以最终只会判断为正确/错误（或者认为50%随机选择），错误率期望为50%

留一法的训练集中数量较多的 label 必然不是留出的样本的 label，必定预测错误，错误率为100%

2. 习题2.4

Solve：

回顾混淆矩阵 confusion matrix

	预测为正	**预测为反**
真实为正	TP	FN
真实为反	FP	TN

真假：对于预测的准确性而言，正反：对于预测的结果而言

真正例率：$TPR = \dfrac{TP}{TP + FN}$, 假正例率：$FPR = \dfrac{FP}{FP+FN}$

查准率：$R = \dfrac{TP}{TP+FN}$, 查全率：$P = \dfrac{TP}{TP+FP}$

就数值而言，$TPR = R$, 其他没有直接的数值关系

3. 习题2.5

Solve：

先回顾 ROC 曲线。我们不妨将 ROC 曲线的横坐标扩大 m^- 倍，纵坐标扩大 m^+ 倍，这样绘制 ROC 曲线图时每一步都走一个单位长度，方便说明.

Step 1. 用分类器对所有数据分类，得到结果为一个 $[0,1]$ 的值，值越大说明越容易被判定为正.

Step 2. 将所有数据按预测结果降序排列

Step 3. 从最大预测结果开始，如果实际为真，向上走一步，如果实际为假，向右走一步

Step 4. 如果有若干样本预测结果相同，先同时向右、上走相应的步数，将起点终点直接相连

下证明 l_{rank} 对应 ROC 曲线上的面积。

对于每一个反样本（即向右走），我们假设没有样本预测结果与之相同。在正样本中：比其预测结果大的，已经在其之前绘制（即已经向上走过了），在 ROC 曲线中表现为曲线之下的部分；反之，比其预测结果小的，还未绘制，且终将在其之后绘制，在 ROC 曲线中表现为曲线之上的部分。曲线之上的部分对应 $\sum_{x^+} \mathbb{I}(f(x^+) < f(x^-))$

假如有 p, q 个正、反样本预测结果相同，除了 $\mathbb{I}(f(x^+) < f(x^-))$ 部分，这对应的是 ROC 曲线纵向方向上对应的矩形，我们还需要计算起点终点直接相连的三角形部分 $\frac{1}{2}pq$. 在 loss 中对应为 $\sum_{x^+} \sum_{x^-} \frac{1}{2}\mathbb{I}(f(x^+) = f(x^-))$.

只需对应缩放，即得到我们的证明。

4. 习题2.9

Solve：

Step 1. 提出原假设和备择假设，离散情况下 $H_0 : P(X = x_i) = p_i, i = 1, 2, \ldots$, 连续情况下 $H_0 : X \sim F(x)$

Step 2. 将 X 的取值范围划分为 k 个互不相交的子区间 $A_1, \ldots, A_k$, 按照习惯，$k \geq 5$.

Step 3. 记落入 A_i 区间的样本个数为 n_i, $\sum_{i=1}^{k} n_i = n$

Step 4. 记随机样本落入 A_i 区间的概率为 q_i

Step 5. 计算

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nq_i)^2}{nq_i}, \chi^2 \sim \chi^2_{k-1}$$