姓名：李千鹏 学号：2021E8014682047 院系：人工智能学院

问题 1.

(1)

贝叶斯风险最小决策规则：$\arg\min_i \left(\sum_{j=1}^{c} \lambda_{ij} P(\omega_j) P\left(x \mid \omega_j\right)\right)$

贝叶斯错误率最小决策规则：$\arg\max_i \left(P\left(\omega_i\right) P\left(x \mid \omega_i\right)\right)$

(2)

当 $\max\left(P\left(\omega_i \mid \mathbf{x}\right) > 1 - \lambda_r/\lambda_s\right)$ 时，分类规则为 $\arg\max_i \left(P\left(\omega_i\right) P\left(x \mid \omega_i\right)\right)$

当 $\max\left(P\left(\omega_i \mid \mathbf{x}\right) \leq 1 - \lambda_r/\lambda_s\right)$ 时，此时拒判，划为 $c+1$ 类

问题 2.

(1)

当 $\mu_1 = \mu_2$ 时，两类同分布，则错误率为 *0.5*

当 $\mu_1 \neq \mu_2$ 时，不妨设 $\mu_1 < \mu_2$

令两类后验概率密度相等即 $0.5 * \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} = 0.5 * \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$，解得 $x_{eq} = \frac{\mu_1+\mu_2}{2}$

由于两个函数关于 x_{eq} 对称，则错误率等于 $P_e = 2 * 0.5 * \int_{x_{eq}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} dx$

令 $\frac{x-\mu_1}{\sigma} = \mu$，带入 P_e 进行换元，可得 $P_e = \frac{1}{\sqrt{2\pi}} \int_{\frac{\mu_2-\mu_1}{2\sigma}}^{+\infty} e^{-\frac{\mu^2}{2}} d\mu$

同样的，当 $\mu_1 > \mu_2$ 时，得 $P_e = \frac{1}{\sqrt{2\pi}} \int_{\frac{\mu_1-\mu_2}{2\sigma}}^{+\infty} e^{-\frac{\mu^2}{2}} d\mu$

当 $\mu_1 = \mu_2$ 时，带入上述式子得 P_e=*0.5*

综上所述，$P_e = \frac{1}{\sqrt{2\pi}} \int_{\frac{|\mu_1-\mu_2|}{2\sigma}}^{+\infty} e^{-\frac{\mu^2}{2}} d\mu$

(2)

当 $a \to +\infty$ 时，式 $\frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}}$，分子趋于 0，分母趋于 $+\infty$，则 $\frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}} \to 0^+$

同时 $P_e \geq 0$，由夹挤原理可得当 $a \to +\infty$ 时，$P_e = 0$

问题 3.

(1)

类条件概率密度函数为：$p(\boldsymbol{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_i}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu_i})\right]$

(2)

类协方差阵不等：$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$

类协方差阵相等：$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i)$

(3)

(1 使用 *PCA* 进行降维

(2 将特征值非常小的数换为一个常数如 *0.000001*，然后用 $\boldsymbol{\phi}\Lambda^{-1}\boldsymbol{\phi}^\top$ 求逆

(3 使用广义逆矩阵代替协方差阵的逆

问题 4.

(1) 推导：

$\boldsymbol{y}$ 均值：$\bar{\boldsymbol{y}} = E(\boldsymbol{y}) = E\left(\boldsymbol{A}^\top\boldsymbol{x}\right) = \boldsymbol{A}^\top\boldsymbol{\mu}$

$\boldsymbol{y}$ 协方差：$\sigma(\boldsymbol{y}) = E\left(\left(\boldsymbol{A}^\top\boldsymbol{x}-\boldsymbol{A}^\top\boldsymbol{\mu}\right)\left(\boldsymbol{A}^\top\boldsymbol{x}-\boldsymbol{A}^\top\boldsymbol{\mu}\right)^\top\right) = E\left(\boldsymbol{A}^\top(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{A}\right) = \boldsymbol{A}^\top\boldsymbol{\Sigma}\boldsymbol{A}$

原样本方差 $\boldsymbol{\Sigma}$ 有等式：$\boldsymbol{\Sigma} = \boldsymbol{\phi}\boldsymbol{\Lambda}\boldsymbol{\phi}^{\boldsymbol{T}}$，$\boldsymbol{\phi}\boldsymbol{\phi}^{\boldsymbol{T}} = \boldsymbol{I}$

带入 $\boldsymbol{y}$ 的协方差得：$\boldsymbol{A}^\top\boldsymbol{\phi}\boldsymbol{\Lambda}\boldsymbol{\phi}^\top\boldsymbol{A} = \boldsymbol{I}$

$\boldsymbol{A} = \boldsymbol{\phi}\boldsymbol{\Lambda}^{-\frac{1}{2}}$ 满足上式，即 $\boldsymbol{A}$ 可满足去相关过程过程

此时 $\boldsymbol{y} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\phi}^{\boldsymbol{T}}\boldsymbol{x}$

再乘以变换矩阵 $\boldsymbol{\phi}$ 得到 *ZCA* 白化结果 $\boldsymbol{\phi}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\phi}^{\boldsymbol{T}}\boldsymbol{x}$

即 $\boldsymbol{A}^{\boldsymbol{T}} = \boldsymbol{\phi}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\phi}^{\boldsymbol{T}} = \boldsymbol{A}$ 为 *ZCA* 的变换矩阵

其中 *ZCA* 的去相关变换矩阵为 $\boldsymbol{\phi}\boldsymbol{\Lambda}^{-\frac{1}{2}}$

(2) 作用：

通过 *ZCA* 白化，使得数据不同维度的特征不相关，同一维度的特征方差为 *1*，实现了去相关过程。

(3)PCA 白化与 *ZCA* 白化作用的异同：

相同点：都实现了去相关过程

不同点：*PCA* 白化对数据进行了降维，*ZCA* 白化没有进行降维。因此 *PCA* 白化处理后的数据，不如 *ZCA* 白化处理后的数据更接近于原始数据。

问题 5.

选取 *MNIST* 标签为 *0*、*1* 的两类数据。*LDF* 的权重通过 *300* 次逻辑斯蒂回归训练得到；*QDF* 的协方差阵的逆则是将特征值非常小的数换为一个常数如 *0.000001*，然后用 $\phi\Lambda^{-1}\phi^{\top}$ 求逆。

最终的分类结果为：*LDF-0.9986,QDF-0.9551*

通过最终的分类结果可以看出 *LDF* 的性能稍高一点，这主要是因为采用了梯度下降的方法对权重进行更新，不断训练最终学习到权重矩阵。*QDF* 的性能稍微低一些，在 *QDF* 中没有进行学习策略，而是基于统计的方法，这就可能会造成分类精度不及 *LDF*。同时类协方差阵不可逆，在对特征值处理后，使用 $\phi\Lambda^{-1}\phi^{\top}$ 计算逆矩阵的方法可能也会对分类结果产生一定的影响。