1 Problem 1

a)
Knowing the general formula for calculating correlation given below, we want to apply two vector to X and Y, find out the maximum value of correlation coefficient. For simplification we set $X = \mathcal{X}$ and $Y = \mathcal{Y}$.
Now we define $\tilde{X} = w_x X$ and $\tilde{Y} = w_y Y$

$$\begin{aligned}
Corr(X,Y) &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \\
&= \frac{Cov(\tilde{X},\tilde{Y})}{\sqrt{Var(\tilde{X})Var(\tilde{Y})}} \\
&= \frac{w_x^\intercal Cov(X,Y) w_y}{\sqrt{w_x^\intercal Var(X) w_x w_y^\intercal Var(Y) w_y}}
\end{aligned}$$

b)
We change above answer st. $Cov(X,Y) = \Sigma_{xy}$,$Var(X) = \Sigma_{xx}$ and $Var(Y) = \Sigma_{yy}$.

$$\rho = \frac{w_x^\intercal \Sigma_{xy} w_y}{\sqrt{w_x^\intercal \Sigma_{xx} w_x w_y^\intercal \Sigma_{yy} w_y}}$$

In order to maximaze the correlation (denominator has no influence on the maximalization, therefore we choose variances) we set:

$$max \rho(w_x, w_y)$$
$$w_x^\intercal \Sigma_x x w_x = 1$$
$$w_y^\intercal \Sigma_y y w_y = 1$$

considering the constraints from a, we get:

$$max w_x^\intercal \Sigma_{xy} w_y$$
$$w_x^\intercal \Sigma_x x w_x = 1$$
$$w_y^\intercal \Sigma_y y w_y = 1$$

We formulate the Lagrange function and calculate the derivatives wrt. w_x and w_y:

$$\mathcal{L}(w_x, w_y, \lambda, \mu) = w_x^\intercal \Sigma_{xy} w_y - \frac{1}{2}\lambda(w_x^\intercal \Sigma_{xx} w_x - 1) - \frac{1}{2}\mu(w_y^\intercal \Sigma_{yy} w_y - 1)$$
$$\frac{\partial \mathcal{L}}{\partial w_x} = \Sigma_{xy} w_y - \lambda \Sigma_{xx} w_x \stackrel{!}{=} 0$$
$$\frac{\partial \mathcal{L}}{\partial w_y} = \Sigma_{xy} w_x - \mu \Sigma_{yy} w_y \stackrel{!}{=} 0$$

Now we multiply $w_x^\intercal$ and $w_y^\intercal$:

$$w_x^\intercal \Sigma_{xy} w_y - \lambda w_x^\intercal \Sigma_{xx} w_x = w_x^\intercal \Sigma_{xy} w_y - \lambda = 0 \leftrightarrow \lambda = w_x^\intercal \Sigma_{xy} w_y$$
$$w_y^\intercal \Sigma_{xy} w_x - \mu w_y^\intercal \Sigma_{yy} w_y = w_y^\intercal \Sigma_{xy} w_x - \mu = 0 \leftrightarrow \mu = w_y^\intercal \Sigma_{xy} w_x$$

Then it follows under the constraints:

$$\lambda = \mu = w_x^\intercal \Sigma_{xy} w_y = \rho(w_x, w_y)$$

Since $\lambda = \mu$,we are able to transform derivatives:

$$\Sigma_{xy} w_y - \lambda \Sigma_{xx} w_x = 0(*)$$
$$\Sigma_{xy} w_x - \lambda \Sigma_{yy} w_y = 0(**)$$

From (**) we get further:

$$\Sigma_{xy} w_x = \lambda \Sigma_{yy} w_y \leftrightarrow w_y = \frac{1}{\lambda} \Sigma_{yy}^{-1} \Sigma_{xy} w_x$$

When we plug w_y into (*), we will receive:

$$\frac{1}{\lambda} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy} w_x - \lambda \Sigma_{xx} w_x = 0$$
$$\leftrightarrow \lambda^2 \Sigma_{xx} w_x = \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy} w_x$$
$$\leftrightarrow \lambda^2 w_x = \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy} w_x$$

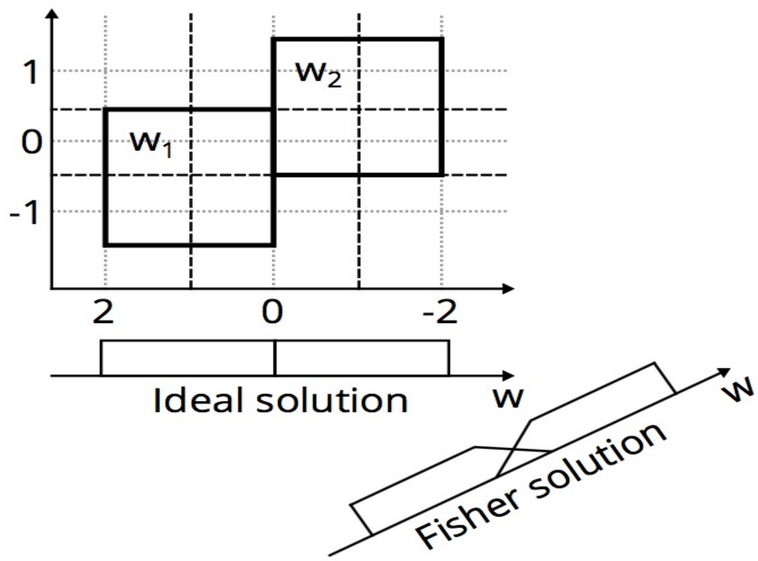
Now we deal with eigenvalue problem:$A w_x = \tilde{\lambda} w_x$, where $\lambda^2 = \tilde{\lambda}$ and $A = \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}$.
To solve w_y it is needed to proceed analogously: $B w_y = \tilde{\lambda} w_y$, where $B = \Sigma_{yy}^{-1} \Sigma_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}$.

c)
From b) we know that $\lambda^2 = \tilde{\lambda}$, which solves the eigenvalues. Since we know that under the constraint $\lambda = w_x^\intercal \Sigma_{xy} w_y = \rho(w_x, w_y)$ we obtain the solution for correlation coefficient at the optimum:

$$\rho(w_x, w_y) = \lambda = \sqrt{\tilde{\lambda}}$$

2 Problem 2

a)



Consider two distributions $p(x|w_1) \sim U(a,b)^2 = U(-1,1)^2$ and $p(x|w_2) \sim U(-1,1)^2$, then we define the expectation as:

$$\mu_1 = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix} \qquad \text{and} \qquad \mu_2 = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$$

We define X as a random variable for first dimension, Y as a random variable for second dimension, and we assume, that two variables are iid. So, we can say, $Cov(X,Y) = Cov(Y,X) = 0$, therefore, we know the variance of X and Y, $Var(X) = Var(Y) = \frac{1}{12}(a-b)^2 = \frac{1}{3}$, The covariance matrix is on the below.

$$\Sigma = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} = \frac{1}{3}I$$

We can compute w^*:

$$w^* = \Sigma_w^{-1}(\mu_2 - \mu_1)\Sigma = 3I\left(\begin{pmatrix} 1 \\ 0.5 \end{pmatrix} - \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}\right) = 3\begin{pmatrix} 2 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

where $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ would be the optimal Bayes solution (from the pic). The linear prediction in Bayes can easily obtained by $p(w_1|x) > p(w_2|x)$, which is different with Fisher discriminant analysis.

b)

When the two classes are generated by two d-dimensional Gaussian distributions,

$$p(x|w_1) = \frac{1}{\sqrt{2\pi det(\Sigma_1)}} exp\left(-\frac{1}{2}(x-\mu_1)^\intercal \Sigma_1^{-1}(x-\mu_1)\right)$$
$$p(x|w_2) = \frac{1}{\sqrt{2\pi det(\Sigma_2)}} exp\left(-\frac{1}{2}(x-\mu_2)^\intercal \Sigma_2^{-1}(x-\mu_2)\right)$$

According to question a, if we decide w_1, therefore,

$$\begin{aligned} p(w_1|x) &> p(w_2|x) \\ \frac{p(x|w_1)p(w_1)}{p(x)} &> \frac{p(x|w_2)p(w_2)}{p(x)} \\ p(x|w_1)p(w_1) &> p(x|w_2)p(w_2) \end{aligned}$$

where we take the logarithm to simplify the computation:

$$\begin{aligned} \ln p(x|w_1) + \ln p(w_1) &> \ln p(x|w_2) + \ln p(w_2) \\ \ln p(x|w_1) - \ln p(x|w_2) &> \ln p(w_2) - \ln p(w_1) \\ -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln det(\Sigma_1) - \frac{1}{2}(x-\mu_1)^\intercal \Sigma_1^{-1}(x-\mu_1) & \\ +\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln det(\Sigma_2) + \frac{1}{2}(x-\mu_2)^\intercal \Sigma_2^{-1}(x-\mu_2) &> \ln p(w_2) - \ln p(w_1) \qquad \text{(PS: plug into above pdf)} \end{aligned}$$

We can identify the first part of the inequality as the mapping $\phi : \mathbb{R}^d \to \mathbb{R}$

$$\phi(x) = (x-\mu_2)^\intercal \Sigma_2^{-1}(x-\mu_2) - (x-\mu_1)^\intercal \Sigma_1^{-1}(x-\mu_1)$$

which is the optimal solutions in Bayes sense.