

Problem Solutions

e-Chapter 9

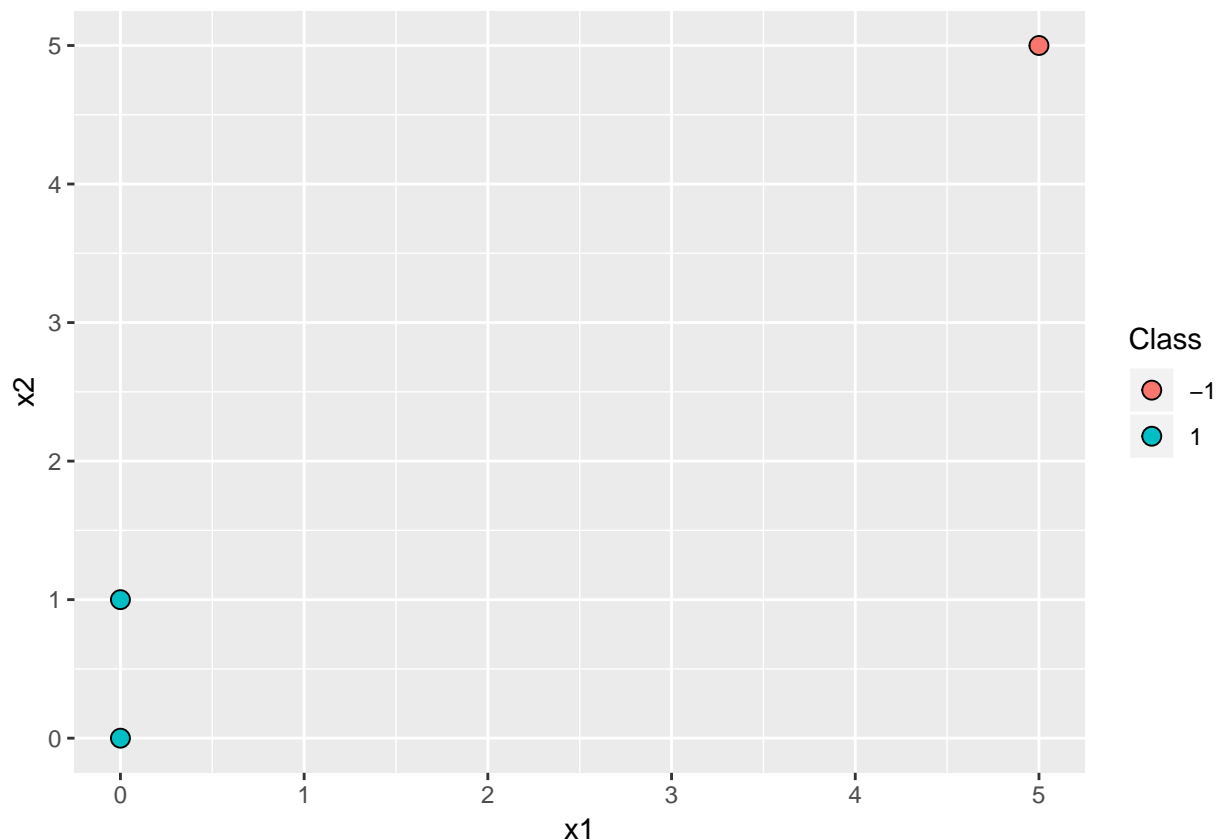
Pierre Paquay

Problem 9.1

(a) We begin by implementing the nearest neighbor method on the raw data.

```
data <- data.frame(x1 = c(0, 0, 5), x2 = c(0, 1, 5))
class <- as.factor(c(1, 1, -1))

ggplot(data, aes(x = x1, y = x2, fill = class)) + geom_point(size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class"))
```



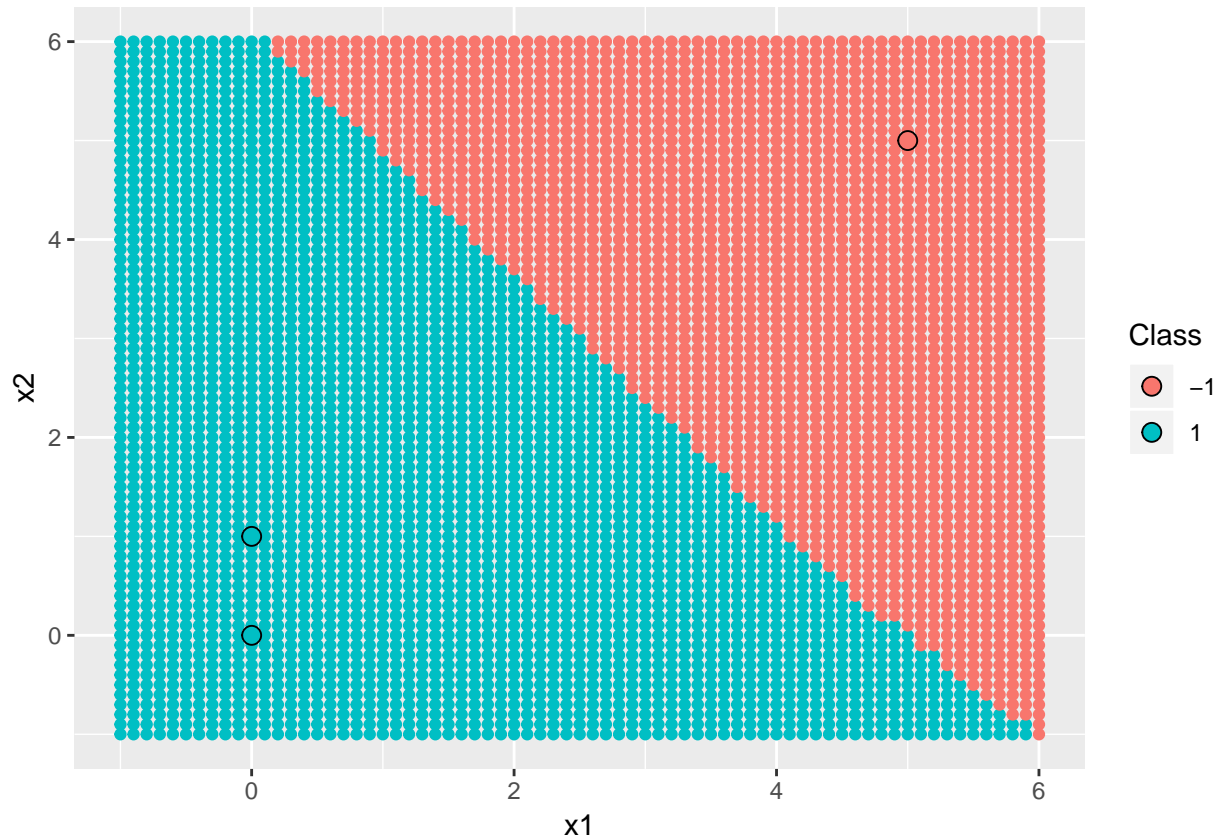
```
grid <- expand.grid(x1 = seq(min(data[, 1] - 1), max(data[, 1] + 1), by = 0.1),
  x2 = seq(min(data[, 2] - 1), max(data[, 2] + 1), by = 0.1))

knn_mod <- knn(data, grid, class, k = 1, prob = TRUE)
```

Below, we show the decision regions of the final hypothesis.

```
ggplot() + geom_point(data = grid, aes(x = x1, y = x2, col = knn_mod)) +
  geom_point(data = data, aes(x = x1, y = x2, fill = class), size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class")) +
```

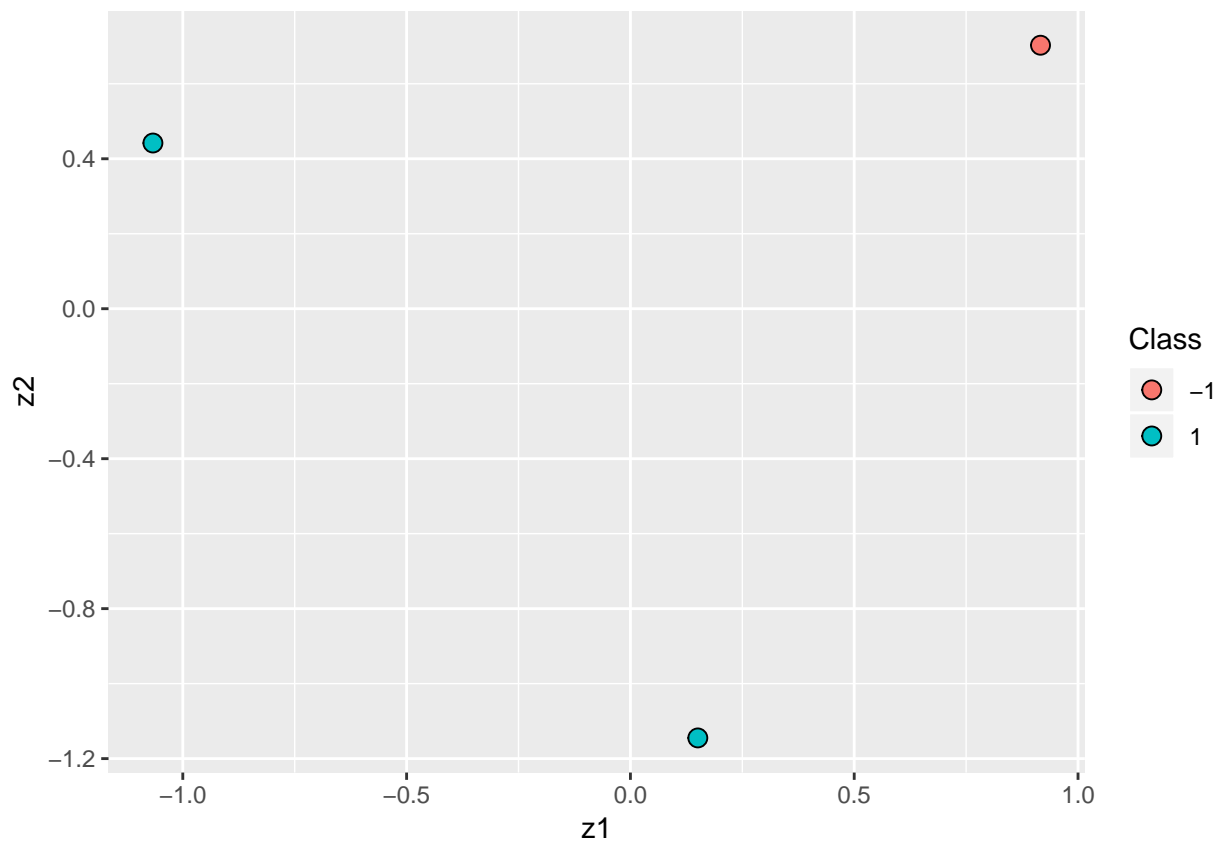
```
guides(col = guide_legend(title = "Class"))
```



(b) Here, we transform to whitened coordinates and we run the nearest neighbor rule.

```
data_centered <- apply(data, 2, function(y) y - mean(y))
sigma <- t(data_centered) %*% as.matrix(data_centered) / 2
sigma_sqr <- sqrtm(sigma)
sigma_sqr_inv <- solve(sigma_sqr)
data_whitened <- as.matrix(data_centered) %*% sigma_sqr_inv
data_whitened <- as.data.frame(data_whitened)
colnames(data_whitened) <- c("z1", "z2")

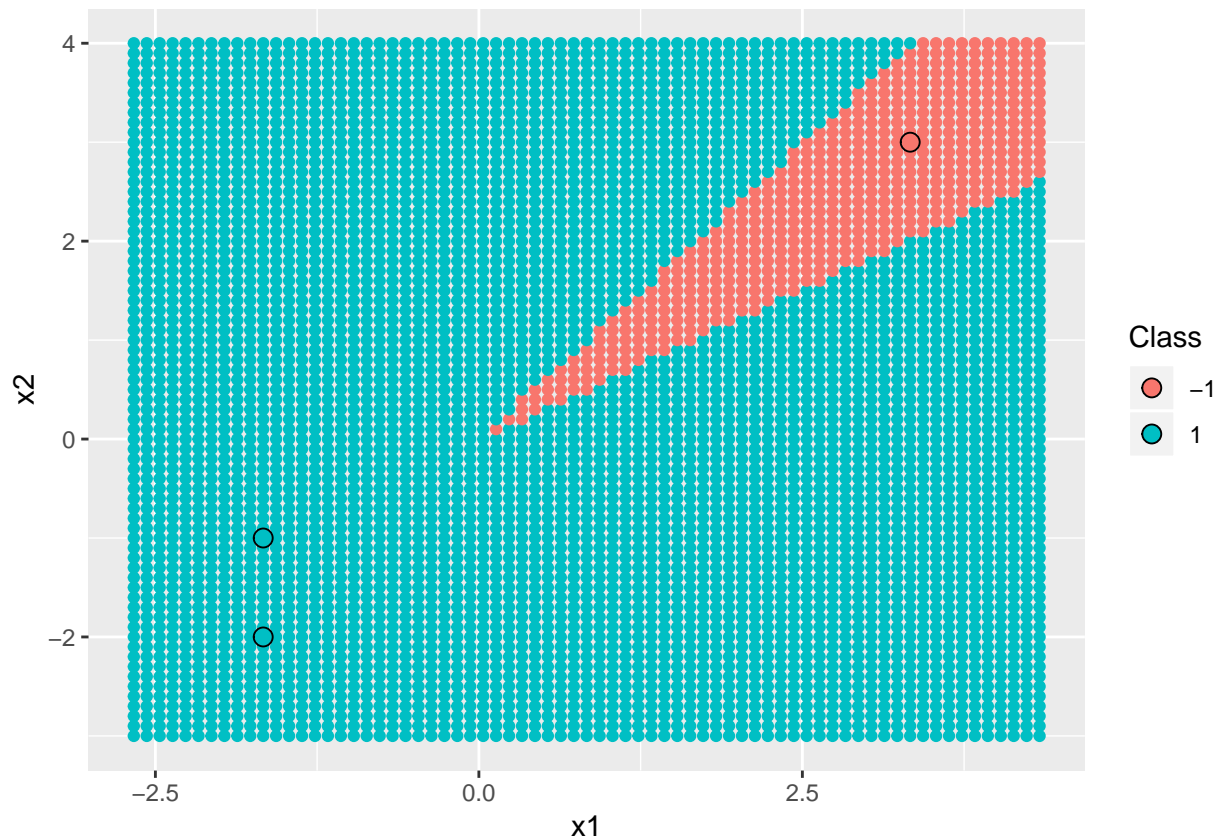
ggplot(data_whitened, aes(x = z1, y = z2, fill = class)) + geom_point(size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class"))
```



```
grid_centered <- expand.grid(x1 = seq(min(data_centered[, 1] - 1),
                                     max(data_centered[, 1] + 1), by = 0.1),
                           x2 = seq(min(data_centered[, 2] - 1),
                                     max(data_centered[, 2] + 1), by = 0.1))
grid_whitened <- as.matrix(grid_centered) %*% sigma_sqr_inv
knn_mod_whitened <- knn(data_whitened, grid_whitened, class, k = 1, prob = TRUE)
```

We show the decision region of the final hypothesis in the original space as well.

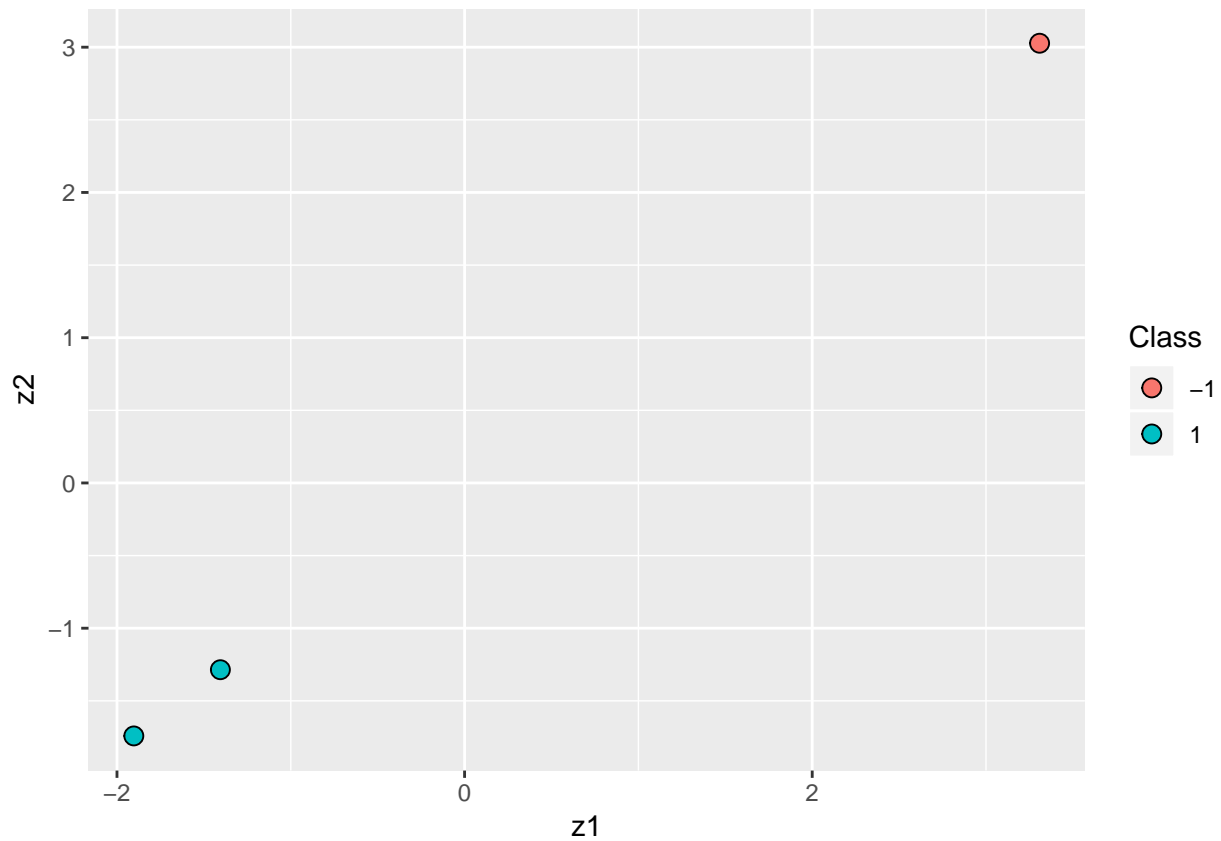
```
ggplot() + geom_point(data = grid_centered, aes(x = x1, y = x2, col = knn_mod_whitened)) +
  geom_point(data = as.data.frame(data_centered), aes(x = x1, y = x2, fill = class),
            size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class")) +
  guides(col = guide_legend(title = "Class"))
```



(c) Finally, we use principal component analysis to reduce the data to 1 dimension for our nearest neighbor classifier.

```
SVD_decomp <- svd(data_centered)
V1 <- SVD_decomp$v[, 1]
Z <- data_centered %*% V1
data_pca <- Z %*% t(V1)
data_pca <- as.data.frame(data_pca)
colnames(data_pca) <- c("z1", "z2")

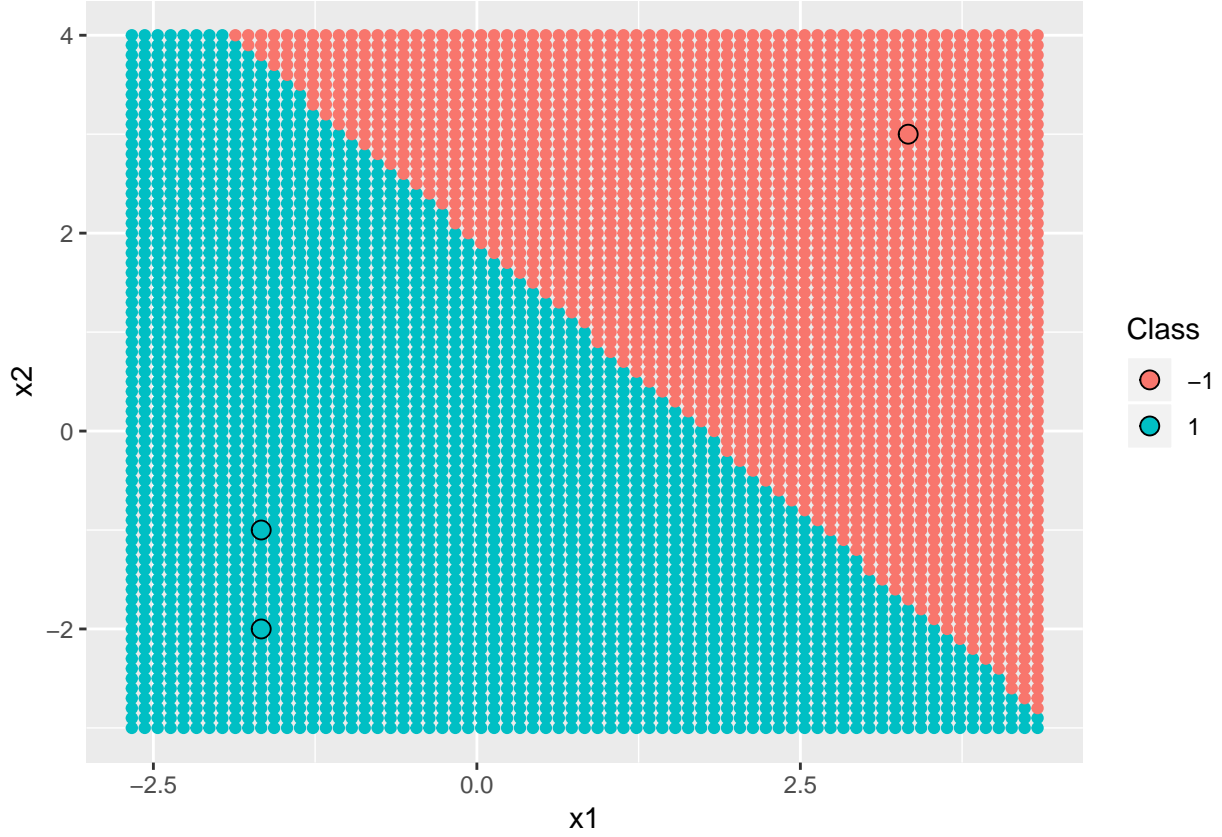
ggplot(data_pca, aes(x = z1, y = z2, fill = class)) + geom_point(size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class"))
```



```
grid_pca <- (as.matrix(grid_centered) %*% V1) %*% t(V1)
knn_mod_pca <- knn(data_pca, grid_pca, class, k = 1, prob = TRUE)
```

Once again, we show the decision regions of the final hypothesis in the original space.

```
ggplot() + geom_point(data = grid_centered, aes(x = x1, y = x2, col = knn_mod_pca)) +
  geom_point(data = as.data.frame(data_centered), aes(x = x1, y = x2, fill = class),
    size = 3, shape = 21) +
  guides(fill = guide_legend(title = "Class")) +
  guides(col = guide_legend(title = "Class"))
```



Problem 9.2

(a) Here, we have $x'_n = x_n + b$. Let us compute \bar{x}' , $\sigma_i'^2$ and Σ' . We have

$$\bar{x}' = \frac{1}{N} \sum_{n=1}^N x'_n = \frac{1}{N} \sum_{n=1}^N (x_n + b) = \bar{x} + b;$$

moreover, we have

$$\sigma_i'^2 = \frac{1}{N} \sum_{n=1}^N (x'_{ni} - \bar{x}'_i)^2 = \frac{1}{N} \sum_{n=1}^N (x_{ni} + b_i - \bar{x}_i - b_i)^2 = \sigma_i^2;$$

and finally

$$\Sigma' = \frac{1}{N} \sum_{n=1}^N (x'_n - \bar{x}')(x'_n - \bar{x}')^T = \frac{1}{N} \sum_{n=1}^N (x_n + b - \bar{x} - b)(x_n + b - \bar{x} - b)^T = \Sigma.$$

- **Centering.** We may write that

$$z'_n = x'_n - \bar{x}' = x_n + b - \bar{x} - b = z_n.$$

- **Normalization.** Here, we have that

$$z'_n = D'(x'_n - \bar{x}')$$

where

$$D' = \text{diag}(1/\sigma'_1, \dots, 1/\sigma'_d) = \text{diag}(1/\sigma_1, \dots, 1/\sigma_d) = D.$$

In that case, we have

$$z'_n = D(x'_n - \bar{x}') = D(x_n + b - \bar{x} - b) = z_n.$$

- **Whitening.** We have

$$z'_n = \Sigma'^{-1/2}(x'_n - \bar{x}') = \Sigma^{-1/2}(x_n + b - \bar{x} - b) = z_n.$$

(b) Here, we have $x'_n = \alpha x_n$ with $\alpha > 0$. Let us compute \bar{x}' , $\sigma_i'^2$ and Σ' . We have

$$\bar{x}' = \frac{1}{N} \sum_{n=1}^N x'_n = \frac{1}{N} \sum_{n=1}^N \alpha x_n = \alpha \bar{x};$$

moreover, we have

$$\sigma_i'^2 = \frac{1}{N} \sum_{n=1}^N (x'_{ni} - \bar{x}'_i)^2 = \frac{1}{N} \sum_{n=1}^N (\alpha x_{ni} - \alpha \bar{x}_i)^2 = \alpha^2 \sigma_i^2;$$

and finally

$$\Sigma' = \frac{1}{N} \sum_{n=1}^N (x'_n - \bar{x}')(x'_n - \bar{x}')^T = \frac{1}{N} \sum_{n=1}^N (\alpha x_n - \alpha \bar{x})(\alpha x_n - \alpha \bar{x})^T = \alpha^2 \Sigma.$$

- **Centering.** We may write that

$$z'_n = x'_n - \bar{x}' = \alpha x_n - \alpha \bar{x} = \alpha z_n \neq z_n.$$

- **Normalization.** Here, we have that

$$z'_n = D'(x'_n - \bar{x}')$$

where

$$D' = \text{diag}(1/\sigma'_1, \dots, 1/\sigma'_d) = \text{diag}(1/(\alpha\sigma_1), \dots, 1/(\alpha\sigma_d)) = \frac{1}{\alpha} D.$$

In that case, we have

$$z'_n = \frac{1}{\alpha} D(x'_n - \bar{x}') = \frac{1}{\alpha} D(\alpha x_n - \alpha \bar{x}) = D(x_n - \bar{x}) = z_n.$$

- **Whitening.** We have

$$z'_n = \Sigma'^{-1/2}(x'_n - \bar{x}') = \frac{1}{\alpha} \Sigma^{-1/2}(\alpha x_n - \alpha \bar{x}) = z_n.$$

(c) Here, we have $x'_n = Ax_n$ with $A = \text{diag}(a_1, \dots, a_d)$ ($a_i \neq 0$). Let us compute \bar{x}' , $\sigma_i'^2$ and Σ' . We have

$$\bar{x}' = \frac{1}{N} \sum_{n=1}^N x'_n = \frac{1}{N} \sum_{n=1}^N Ax_n = A\bar{x};$$

moreover, we have

$$\sigma_i'^2 = \frac{1}{N} \sum_{n=1}^N (x'_{ni} - \bar{x}'_i)^2 = \frac{1}{N} \sum_{n=1}^N (a_i x_{ni} - a_i \bar{x}_i)^2 = a_i^2 \sigma_i^2;$$

and finally

$$\Sigma' = \frac{1}{N} \sum_{n=1}^N (x'_n - \bar{x}')(x'_n - \bar{x}')^T = \frac{1}{N} \sum_{n=1}^N (Ax_n - A\bar{x})(Ax_n - A\bar{x})^T = A\Sigma A^T.$$

- **Centering.** We may write that

$$z'_n = x'_n - \bar{x}' = Ax_n - A\bar{x} = Az_n \neq z_n.$$

- **Normalization.** Here, we have that

$$z'_n = D'(x'_n - \bar{x}')$$

where

$$D' = \text{diag}(1/\sigma'_1, \dots, 1/\sigma'_d) = \text{diag}(1/(a_1\sigma_1), \dots, 1/(a_d\sigma_d));$$

which means that $D'A = D$. In that case, we have

$$z'_n = D'(x'_n - \bar{x}') = D'A(x_n - \bar{x}) = D(x_n - \bar{x}) = z_n.$$

- **Whitening.** We have

$$z'_n = \Sigma'^{-1/2}(x'_n - \bar{x}') = (A\Sigma A^T)^{-1/2}A(x_n - \bar{x}) \neq z_n.$$

(d) Here, we have $x'_n = Ax_n$ with $\det(A) \neq 0$. Let us compute \bar{x}' , $\sigma_i'^2$ and Σ' . We have

$$\bar{x}' = \frac{1}{N} \sum_{n=1}^N x'_n = \frac{1}{N} \sum_{n=1}^N Ax_n = A\bar{x};$$

moreover, we have

$$\sigma_i'^2 = \frac{1}{N} \sum_{n=1}^N (x'_{ni} - \bar{x}'_i)^2 = \frac{1}{N} \sum_{n=1}^N (l_i^T x_n - l_i^T \bar{x})^2 = \frac{1}{N} \sum_{n=1}^N l_i^T (x_n - \bar{x})(x_n - \bar{x})^T l_i = l_i^T \Sigma l_i$$

where l_i is the i th row of A ; and finally

$$\Sigma' = \frac{1}{N} \sum_{n=1}^N (x'_n - \bar{x}')(x'_n - \bar{x}')^T = \frac{1}{N} \sum_{n=1}^N (Ax_n - A\bar{x})(Ax_n - A\bar{x})^T = A\Sigma A^T.$$

- **Centering.** We may write that

$$z'_n = x'_n - \bar{x}' = Ax_n - A\bar{x} = Az_n \neq z_n.$$

- **Normalization.** Here, we have that

$$z'_n = D'(x'_n - \bar{x}')$$

where

$$D' = \text{diag}(1/\sigma'_1, \dots, 1/\sigma'_d) = \text{diag}(1/(l_1^T \Sigma l_1), \dots, 1/(l_d^T \Sigma l_d)).$$

In that case, we have

$$z'_n = D'(x'_n - \bar{x}') = D'A(x_n - \bar{x}) = D'Az_n \neq z_n.$$

- **Whitening.** We have

$$z'_n = \Sigma'^{-1/2}(x'_n - \bar{x}') = (A\Sigma A^T)^{-1/2}A(x_n - \bar{x}) \neq z_n.$$

Problem 9.3

We may write

$$\Gamma = \text{diag}(\lambda_1, \dots, \lambda_d)$$

where $\lambda_i > 0$. With this in mind, it is easy to describe $\Gamma^{1/2}$ and $\Gamma^{-1/2}$; we have

$$\Gamma^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}) \text{ and } \Gamma^{-1/2} = \text{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_d}).$$

To prove the first fact, we simply need to compute $\Gamma^{1/2}\Gamma^{1/2}$, we have

$$\Gamma^{1/2}\Gamma^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})\text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}) = \text{diag}(\lambda_1, \dots, \lambda_d) = \Gamma;$$

to prove the second fact, we need to compute $\Gamma^{1/2}\Gamma^{-1/2}$ (the same reasoning can be applied to $\Gamma^{-1/2}\Gamma^{1/2}$), we have

$$\Gamma^{1/2}\Gamma^{-1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})\text{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_d}) = I.$$

Now, to prove that $\Sigma^{1/2} = U\Gamma^{1/2}U^T$, we write that

$$\Sigma^{1/2}\Sigma^{1/2} = U\Gamma^{1/2}\underbrace{U^T U}_{=I}\Gamma^{1/2}U^T = U\Gamma U^T = \Sigma.$$

Then, to prove that $\Sigma^{-1/2} = U\Gamma^{-1/2}U^T$, we write that

$$\Sigma^{-1/2}\Sigma^{1/2} = U\Gamma^{-1/2}\underbrace{U^T U}_{=I}\Gamma^{1/2}U^T = UIU^T = I.$$

Problem 9.4

We know that $A = V\psi$, consequently we get

$$V^T A = \underbrace{V^T V}_{=I} \psi = \psi$$

because V is an orthogonal matrix. Moreover, we also have that

$$\psi^T \psi = A^T \underbrace{V V^T}_{=I} A = A^T A = I$$

since A is an orthogonal matrix as well; this means that ψ is also an orthogonal matrix.