Problem 1

Higher degree polynomial functions may cause overfitting. Every sample data contains certain error, though using higher degree functions may fit every known training points, it is impossible to fit all upcoming test points. Though using quadratic error functions may not fit all training points, it can be well enough to fit most training and test points.

Problem 2

a. Get mean value for each feature:

$$U_1 = (-2-5-3+0-8-2+1+5-1+6) / 10 = -0.9$$
$$U_2 = (1-4+1+3+11+5+0-1-3+1) / 10 = 1.4$$

Center the points:

Data: { [-1.1 -0.4], [-4.1 -5.4], [-2.1 -0.4], [0.9 1.6], [-7.1 9.6], [-1.1 3.6], [1.9 -1.4], [5.9 -2.4], [-0.1 -4.4], [6.9 -0.4] }

Compute covariance matrix:

$$\sum(X) = \frac{X^T X}{N-1} = \begin{bmatrix} 17.8778 & -7.3778 \\ -7.3778 & 18.2667 \end{bmatrix}$$

Compute eigenvalues and vectors:

$$\lambda^2 - \lambda(17.8778 + 18.2667) + (17.8778 \times 18.2667 - 7.3778^2) = 0$$
$$\lambda_1 = 10.6919, \lambda_2 = 25.4526$$

Values: [10.6919, 25.4526]

$$(A - \lambda I)x = \begin{bmatrix} 7.1859 & -7.3778 \\ -7.3778 & -7.1859 \end{bmatrix} x = 0$$

2 eigenvectors before normalization: $[-7.3778\ -7.1859]^T$, $[7.1859\ -7.3778]^T$

Vectors: $[-0.7164, -0.6977]^T$, $[-0.6977, 0.7164]^T$

b. Best eigenvector is: $[-0.6977, 0.7164]^T$, eigenvalue: 25.4526

$P_1 = [-1.1\ -0.4] * [-0.6977, 0.7164]^T = 0.4809$

$P_2 = [-4.1\ -5.4] * [-0.6977, 0.7164]^T = -1.0080$

$P_3 = [-2.1\ -0.4] * [-0.6977, 0.7164]^T = 1.1786$

$P_4 = [0.9\ 1.6] * [-0.6977, 0.7164]^T = = 0.5183$

$P_5 = [-7.1\ 9.6] * [-0.6977, 0.7164]^T = 11.8311$

$P_6 = [-1.1\ 3.6] * [-0.6977, 0.7164]^T = 3.3465$

$P_7 = [1.9\ -1.4] * [-0.6977, 0.7164]^T = -2.3286$

$P_8 = [5.9\ -2.4] * [-0.6977, 0.7164]^T = -5.8358$

$P_9 = [-0.1\ -4.4] * [-0.6977, 0.7164]^T = -3.0824$

$P_{10} = [6.9\ -0.4] * [-0.6977, 0.7164]^T = -5.1007$

Problem 3

a. Positive: { [-2 1], [-5 -4], [-3 1], [0 3], [-8 11] } Negative: { [-2 5], [1 0], [5 -1], [-1 -3], [6 1] }

Feature 1:

There are 5 positive and 5 negative samples respectively. It has 9 distinct values in total. There are one positive and negative sample each for -2, for other values, there is only one positive or negative.

$$IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A)$$

$$= I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \sum_{i=1}^{k} \frac{p_i+n_i}{p_i+n_i} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$= I\left(\frac{1}{2}, \frac{1}{2}\right) - \frac{4}{10} I(1,0) - \frac{4}{10} I(0,1) - \frac{1}{5} I\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{5} = 0.8$$

Feature 2:

There are 5 positive and 5 negative samples respectively. It has 8 distinct values in total. There are one positive and two negative samples for -1, for other values, there is only one positive or negative.

$$IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A)$$

$$= I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \sum_{i=1}^{k} \frac{p_i+n_i}{p_i+n_i} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$= I\left(\frac{1}{2}, \frac{1}{2}\right) - \frac{4}{10} I(1,0) - \frac{3}{10} I(0,1) - \frac{3}{10} I\left(\frac{1}{3}, \frac{2}{3}\right)$$

$$= 1 - \frac{3}{10} \times \left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right) = 0.7245$$

b. According to last part, feature 1 is more discriminating.

c. Means for each class:

$$\mu_1 = [-3.6, 2.4]^T$$
$$\mu_2 = [1.8, 0.4]^T$$

Scatter matrix:

$$S_1 = (N_1 - 1) \times \mathrm{cov}(C_1) = 4 \times \begin{bmatrix} 9.3 & -7.45 \\ -7.45 & 29.8 \end{bmatrix} = \begin{bmatrix} 37.2 & -29.8 \\ -29.8 & 119.2 \end{bmatrix}$$

$$S_2 = (N_2 - 1) \times \mathrm{cov}(C_2) = 4 \times \begin{bmatrix} 12.7 & -2.4 \\ -2.4 & 8.8 \end{bmatrix} = \begin{bmatrix} 50.8 & -9.6 \\ -9.6 & 35.2 \end{bmatrix}$$

Within class scatter:

$$S_w = S_1 + S_2 = \begin{bmatrix} 88 & -39.4 \\ -39.4 & 154.4 \end{bmatrix}$$

$${S_w}^{-1} = \begin{bmatrix} 0.0128 & 0.0033 \\ 0.0033 & 0.0073 \end{bmatrix}$$

Eigenvectors:

$$v = {S_w}^{-1}(\mu_1 - \mu_2) = [-0.0627 \quad -0.0031]^T$$

d. Projection:

Class 1:

P_1 = [-2 1] * [-0.0627 -0.0031]^T = 0.1223

P_2 = [-5 -4] * [-0.0627 -0.0031]^T = 0.3259

P_3 = [-3 1] * [-0.0627 -0.0031]^T = 0.1850

P_4 = [0 3] * [-0.0627 -0.0031]^T = -0.0093
P_5 = [-8 11] * [-0.0627 -0.0031]^T = 0.4675
Class 2:
P_1 = [-2 5] * [-0.0627 -0.0031]^T = 0.1099
P_2 = [1 0] * [-0.0627 -0.0031]^T = -0.0627
P_3 = [5 -1] * [-0.0627 -0.0031]^T = -0.3104
P_4 = [-1 -3] * [-0.0627 -0.0031]^T = 0.0720
P_5 = [6 1] * [-0.0627 -0.0031]^T = -0.3793

e. I think it is acceptable but not good. P_4 in class 1 is projected into negative part, which is supposed to contain class 2. P_1 and P_4 in class 2 are projected to positive part, however they are not mixed with class 1 points. Not all points in each class are separated, but in general over half of the points are discriminated clearly.

Problem 4

Since 2nd value is dependent, we can separate this matrix:

$$X = [-2 \quad -5 \quad -3 \quad 0 \quad -8 \quad -2 \quad 1 \quad 5 \quad -1 \quad 6]^T$$
$$Y = [1 \quad -4 \quad 1 \quad 3 \quad 11 \quad 5 \quad 0 \quad -1 \quad -3 \quad 1]^T$$

A value of β_1 needs to be determined so that $\beta_1 x$ has the least square error.
Formula for square error is:

$$Error(\beta_1) = \sum_k (y^{(k)} - \beta_1 x^{(k)})^2$$

When the derivative of Error(β_1) equals 0, that value is the best choice for β_1.
That is:

$$\beta_1 = \frac{\sum_k y^{(k)} x^{(k)}}{\sum_k (x^{(k)})^2} = (\sum_k \left(x^{(k)}\right)^2)^{-1} \sum_k y^{(k)} x^{(k)}$$

Since $\sum_k \left(x^{(k)}\right)^2 = X^T X$ and $\sum_k y^{(k)} x^{(k)} = X^T Y$, so that $\beta_1 = (X^T X)^{-1} X^T Y$

$\sum_k \left(x^{(k)}\right)^2 = 169$, $\sum_k y^{(k)} x^{(k)} = -79$

So $\beta_1 = \frac{-79}{169} = -0.4675$, and $\hat{y} = \beta_1 x = -0.4675x$