Decision Models - Assignment 2

Matteo Suardi

May 2025

Exercise 1

Let

$$f(x, y) = 2x^2 + y^2 - xy + x - y$$

Answer the following:

(a) Find the stationary points.

(b) Classify each stationary point.

(c) Perform one iteration of the gradient method with exact line search, starting from $(0, 1)$.

(d) Perform one iteration of Newton's method, starting from $(0, 1)$.

(a) Stationary points

We compute the gradient:

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 4x - y + 1 \\ 2y - x - 1 \end{bmatrix}$$

Set $\nabla f(x, y) = 0$, we solve:

$$\begin{cases} 4x - y + 1 = 0 \\ 2y - x - 1 = 0 \end{cases} \Rightarrow x = -\frac{1}{7}, \quad y = \frac{3}{7}$$

Thus, the only stationary point is $\left(-\frac{1}{7}, \frac{3}{7}\right)$.

(b) Classification via Hessian

Compute second derivatives to obtain the Hessian matrix:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$$

We compute the determinant: $\det(H) = 4 \cdot 2 - (-1)^2 = 8 - 1 = 7 > 0$
Also, $\frac{\partial^2 f}{\partial x^2} = 4 > 0$
Conclusion: H is positive definite $\rightarrow$ the point is a **local minimum**.

(c) Gradient method - one iteration

Let $z_0 = (0, 1)$. Compute gradient in z_0:

$$\nabla f(0,1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow d_0 = -\nabla f(0,1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

To do the line search exactly, we set $z(t) = (0, 1-t)$ as the result of the computation:

$$z(t) = z_0 + t \cdot d_0 = (0,1) + t(0,-1) = (0, 1-t).$$

Then:

$$f(0, 1-t) = (1-t)^2 - (1-t) = t^2 - t$$

Minimize $f(t)$:

$$\min_t f(t) = \min_t (t^2 - t) \Rightarrow t_0 = \frac{1}{2}$$

Update:

$$z_1 = (0, 1 - \frac{1}{2}) = \left(0, \frac{1}{2}\right)$$

(d) Newton's method - one iteration

Start from $z_0 = (0, 1)$

$$\nabla f(0,1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$$

Compute Newton direction:

$$d = -H^{-1}\nabla f(0,1) = -\frac{1}{7}\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{7}\begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} -1/7 \\ -4/7 \end{bmatrix}$$

Then:

$$z_1 = (0,1) + \left(-\frac{1}{7}, -\frac{4}{7}\right) = \left(-\frac{1}{7}, \frac{3}{7}\right)$$

Conclusion: One Newton step lands exactly at the stationary point found in (a).

Exercise 2

We consider the problem:

$$\min_{x,y} \quad f(x,y) = x^2 + y^2 - 2y$$

subject to:

$$\begin{cases} g_1(x,y) = x^2 + y^2 - 1 \leq 0 \\ g_2(x,y) = -y \leq 0 \end{cases}$$

(a) KKT conditions

We define the Lagrangian:

$$\mathcal{L}(x,y,\lambda_1,\lambda_2) = x^2 + y^2 - 2y + \lambda_1(x^2 + y^2 - 1) + \lambda_2(-y)$$

Compute the gradient of the Lagrangian:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x(1+\lambda_1), \qquad \frac{\partial \mathcal{L}}{\partial y} = 2y(1+\lambda_1) - 2 - \lambda_2$$

The KKT conditions are:

(Stationarity) $\begin{cases} 2x(1+\lambda_1) = 0 \\ 2y(1+\lambda_1) - 2 - \lambda_2 = 0 \end{cases}$

(Primal feasibility) $\begin{cases} x^2 + y^2 - 1 \leq 0 \\ -y \leq 0 \end{cases}$

(Dual feasibility) $\lambda_1 \geq 0, \quad \lambda_2 \geq 0$

(Complementarity slackness) $\begin{cases} \lambda_1(x^2 + y^2 - 1) = 0 \\ \lambda_2(-y) = 0 \end{cases}$

(b) Case analysis

We consider all combinations of possible values of λ_1 and λ_2, using the complementarity conditions:

- If $\lambda_i > 0$, then the corresponding constraint must be **active** (i.e., equality).
- If the constraint is **not active** (strict inequality), then $\lambda_i = 0$.

Case 1: $\lambda_1 = 0,\ \lambda_2 = 0$

From stationarity:

$2x = 0 \Rightarrow x = 0$

$2y - 2 = 0 \Rightarrow y = 1$

Check constraints:

$x^2 + y^2 = 1 \Rightarrow \text{active} \quad (\text{ok, even if } \lambda_1 = 0)$

$y = 1 > 0 \Rightarrow \text{inactive} \quad \Rightarrow \lambda_2 = 0 \text{ (ok)}$

Valid KKT point.

Case 2: $\lambda_1 > 0,\ \lambda_2 = 0$

$$\begin{aligned}
&\text{Complementarity: } x^2 + y^2 = 1 \quad \text{(active)}\\
&\text{Stationarity:}\\
&\quad 2x(1+\lambda_1) = 0 \Rightarrow x = 0 \quad (\text{since } \lambda_1 > 0)\\
&\quad 2y(1+\lambda_1) - 2 = 0 \Rightarrow y = \frac{1}{1+\lambda_1} \in (0,1)\\
&\text{Check feasibility: } x^2 + y^2 = y^2 = \frac{1}{(1+\lambda_1)^2} \le 1 \quad \to \text{implies } \lambda_1 = 0 \text{ (contradiction)}\\
&\textbf{Rejected: violates complementarity for } \lambda_1 > 0.
\end{aligned}$$

Case 3: $\lambda_1 = 0,\ \lambda_2 > 0$

$$\begin{aligned}
&\text{Complementarity: } y = 0\\
&\text{Stationarity:}\\
&\quad 2x = 0 \Rightarrow x = 0\\
&\quad -2 - \lambda_2 = 0 \Rightarrow \lambda_2 = -2 < 0\\
&\textbf{Rejected: violates dual feasibility.}
\end{aligned}$$

Case 4: $\lambda_1 > 0,\ \lambda_2 > 0$

$$\begin{aligned}
&\text{Complementarity: } \begin{cases} x^2 + y^2 = 1 \\ y = 0 \end{cases} \Rightarrow x^2 = 1 \Rightarrow x = \pm 1\\
&\text{Stationarity:}\\
&\quad 2x(1+\lambda_1) = 0 \Rightarrow x = 0 \quad \text{contradicts } x = \pm 1\\
&\textbf{Rejected: incompatible with stationarity.}
\end{aligned}$$

Conclusion: The only feasible solution satisfying all KKT conditions is:

$$(x^*, y^*) = (0,1), \quad \lambda_1 = 0, \quad \lambda_2 = 0$$

(c) Optimal solution and value

The only valid KKT point is:

$$(x^*, y^*) = (0,1), \quad \lambda_1^* = 0, \quad \lambda_2^* = 0$$

Objective function value:

$$f(0,1) = 0^2 + 1^2 - 2 \cdot 1 = -1$$

Conclusion: The optimal solution is $(x^*, y^*) = (0,1)$ with minimum value $f^* = -1$.

Exercise 3

Theoretical formulation

Primal problem: Given training data $(x_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}$, the primal soft-margin SVM solves:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell} \xi_i \\ \text{subject to:} \quad & y_i(w^\top \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \quad \text{for all } i = 1, \ldots, \ell \end{aligned}$$

where $\phi(\cdot)$ maps data into a higher-dimensional feature space.

Dual problem: By duality, the optimization becomes:

$$\begin{aligned} \max_{\lambda} \quad & \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j K(x_i, x_j) \\ \text{subject to:} \quad & \sum_{i=1}^{\ell} \lambda_i y_i = 0, \quad 0 \leq \lambda_i \leq C \end{aligned}$$

where $K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^2$ is the polynomial kernel of degree 2.

Decision function: Once the optimal multipliers λ^* are found, the decision function for a new input x is:

$$f(x) = \text{sign}\left(\sum_{i=1}^{\ell} \lambda_i^* y_i K(x_i, x) + b^* \right)$$

Bias b^* computation: To compute the bias b^*, we use the KKT condition on a support vector x_k such that $0 < \lambda_k < C$:

$$b^* = y_k - \sum_{i=1}^{\ell} \lambda_i y_i K(x_i, x_k)$$

Alternatively, for better stability, we average over all such support vectors:

$$b^* = \frac{1}{|M|} \sum_{k \in M} \left(y_k - \sum_{i=1}^{\ell} \lambda_i y_i K(x_i, x_k) \right)$$

where $M = \{i \in \{1, \ldots, \ell\} : 0 < \lambda_i < C\}$.

Support Vectors: Support vectors are those training points for which $\lambda_i > 0$. These are the only points contributing to the decision boundary.

(a) Dual optimization model in AMPL

The dual of the soft-margin SVM problem with polynomial kernel is formulated as:

```
set N;
param y {N};
param C;
param K {N,N};

var lambda {i in N} >= 0, <= C;

maximize DualObjective:
    sum {i in N} lambda[i]
    - 0.5 * sum {i in N, j in N} lambda[i] * lambda[j] * y[i] * y[j] * K[i,j];

subject to Balance:
    sum {i in N} lambda[i] * y[i] = 0;
```

Using the following AMPL script, we solved the dual problem with CPLEX:

```
reset;
model svm.mod;
data svm_poly_kernel.dat;
option solver cplex;
solve;
display lambda;
```

Result: Optimal objective value: $\boxed{11.905}$, found in 0 simplex iterations.

(b) Optimal λ^* and support vectors

To extract the support vectors, we filtered the nonzero λ_i values:

```
display {i in N: lambda[i] > 1e-5} lambda[i];
```

Nonzero λ_i:

$$\begin{aligned}
&\lambda_1 = 1, \quad \lambda_6 = 0.1461, \quad \lambda_7 = 1, \quad \lambda_9 = 1, \quad \lambda_{13} = 1, \quad \lambda_{15} = 1, \\
&\lambda_{16} = 1, \quad \lambda_{18} = 1, \quad \lambda_{19} = 0.1815, \quad \lambda_{22} = 0.3005, \quad \lambda_{26} = 1, \\
&\lambda_{27} = 1, \quad \lambda_{29} = 0.5501, \quad \lambda_{30} = 1, \quad \lambda_{33} = 1, \quad \lambda_{34} = 0.4770, \\
&\lambda_{36} = 1, \quad \lambda_{39} = 1
\end{aligned}$$

Support vectors:

$$\mathrm{SV} = \{1,\ 6,\ 7,\ 9,\ 13,\ 15,\ 16,\ 18,\ 19,\ 22,\ 26,\ 27,\ 29,\ 30,\ 33,\ 34,\ 36,\ 39\}$$

(c) Scalar b^* in the decision function

We estimate b^* using the KKT condition on a support vector k such that $0 < \lambda_k < C$. Since we are using the kernel matrix K, we compute:

```
param b_star;
let b_star := y[6] - sum {i in N} lambda[i] * y[i] * K[i,6];
display b_star;
```

This gives $b^* = \boxed{-6.98115}$. To obtain a more stable value, we average over all support vectors not on the margin:

```
set MARGIN := {i in N: lambda[i] > 1e-5 && lambda[i] < C - 1e-5};

param b_avg;
let b_avg := (1 / card(MARGIN)) * sum {k in MARGIN} (
    y[k] - sum {i in N} lambda[i] * y[i] * K[i,k]
);
display b_avg;
```

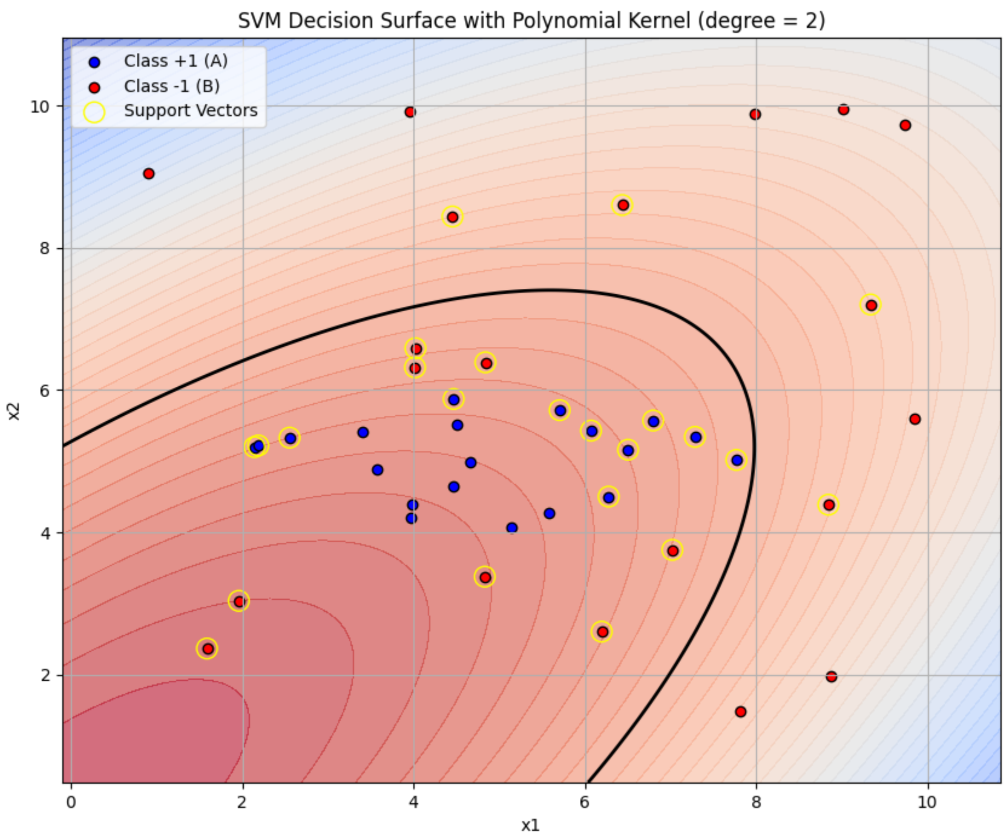
The result is: $b^* = \boxed{-5.87801}$

(d) Plot of the separating surface

To visualize the separating surface found by the nonlinear SVM with polynomial kernel of degree $p = 2$, we used the following Python script. It employs `scikit-learn`'s `SVC` model with the same parameters as in AMPL:

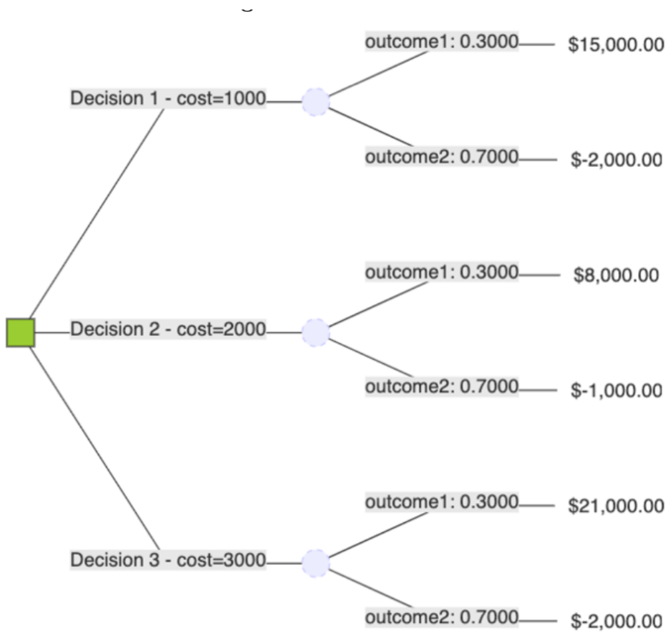
```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import cm
from sklearn.svm import SVC

# Load data from files
A = np.loadtxt("A.txt")
B = np.loadtxt("B.txt")

# Stack data and assign labels
X = np.vstack((A, B))
y = np.concatenate((np.ones(len(A)), -np.ones(len(B))))

# Train nonlinear SVM with polynomial kernel (degree=2)
clf = SVC(C=1, kernel='poly', degree=2)
clf.fit(X, y)

# Create grid to evaluate decision function
xx, yy = np.meshgrid(np.linspace(X[:, 0].min() - 1, X[:, 0].max() + 1, 500),
                     np.linspace(X[:, 1].min() - 1, X[:, 1].max() + 1, 500))
grid = np.c_[xx.ravel(), yy.ravel()]
Z = clf.decision_function(grid).reshape(xx.shape)
```

```
# Plot the decision boundary and margins
plt.figure(figsize=(10, 8))
plt.contourf(xx, yy, Z, levels=50, cmap=cm.coolwarm, alpha=0.6)
plt.contour(xx, yy, Z, levels=[0], colors='k', linewidths=2)

# Plot the original data
plt.scatter(A[:, 0], A[:, 1], color='blue', label='Class +1 (A)', edgecolors='k')
plt.scatter(B[:, 0], B[:, 1], color='red', label='Class -1 (B)', edgecolors='k')

# Highlight support vectors
plt.scatter(clf.support_vectors_[:, 0], clf.support_vectors_[:, 1],
            s=150, facecolors='none', edgecolors='yellow', label='Support Vectors')

plt.legend()
plt.title("SVM Decision Surface with Polynomial Kernel (degree = 2)")
plt.xlabel("x1")
plt.ylabel("x2")
plt.grid(True)
plt.show()
```



The plot shows the the support vectors and the class distributions in the input space.

Exercise 4



(a) Optimal strategy

$$\text{Decision 1:} \quad \begin{aligned} EV_1 &= (15{,}000 - 1{,}000) \cdot 0.3 + (-2{,}000 - 1{,}000) \cdot 0.7 \\ &= 14{,}000 \cdot 0.3 + (-3{,}000) \cdot 0.7 \\ &= 4{,}200 - 2{,}100 = \boxed{2{,}100} \end{aligned}$$

$$\text{Decision 2:} \quad \begin{aligned} EV_2 &= (8{,}000 - 2{,}000) \cdot 0.3 + (-1{,}000 - 2{,}000) \cdot 0.7 \\ &= 6{,}000 \cdot 0.3 + (-3{,}000) \cdot 0.7 \\ &= 1{,}800 - 2{,}100 = \boxed{-300} \end{aligned}$$

$$\text{Decision 3:} \quad \begin{aligned} EV_3 &= (21{,}000 - 3{,}000) \cdot 0.3 + (-2{,}000 - 3{,}000) \cdot 0.7 \\ &= 18{,}000 \cdot 0.3 + (-5{,}000) \cdot 0.7 \\ &= 5{,}400 - 3{,}500 = \boxed{1{,}900} \end{aligned}$$

The optimal strategy is the first one.

(b) Value of perfect information (VPI)

We now compute the expected value of perfect information, i.e., the expected gain if we could know in advance whether outcome 1 or outcome 2 will occur.

For each possible outcome, we choose the best decision available.

Outcome 1 (probability = 0.3):

- Best decision: Decision 3
- Net gain: $21{,}000 - 3{,}000 = 18{,}000$
- Weighted value: $0.3 \cdot 18{,}000 = 5{,}400$

Outcome 2 (probability = 0.7):

- Best decision: Decision 2
- Net gain: $-1{,}000 - 2{,}000 = -3{,}000$
- Weighted value: $0.7 \cdot (-3{,}000) = -2{,}100$

Expected value with perfect information:

$$\text{EV}_{\text{with perfect info}} = 5{,}400 - 2{,}100 = \boxed{3{,}300}$$

Expected value without information:
From part (a), the best expected value among all decisions is:

$$\text{EV}_{\text{without info}} = \max\{2{,}100,\ -300,\ 1{,}900\} = \boxed{2{,}100}$$

Value of perfect information (VPI):

$$\text{VPI} = 3{,}300 - 2{,}100 = \boxed{1{,}200}$$

Conclusion: The maximum amount we should be willing to pay for perfect information is \$1,200.