Perceptron

Machine Learning Techniques

References and Credits

- The content presented in these slides is derived from professor Arun Rajkumar's lectures and slides in the MLT course. This is the "ground truth" for almost all the content in these slides. As a result, these slides should be viewed as a presentation of the same content in the professor's lectures using a different medium. At the same time, these slides are not meant to be a replacement for the lectures.
- The method of incrementally displaying content on slides is borrowed from professor Mitesh Khapra.
- These slides were prepared using the tool mathcha.io.

Linear Separability

$$P(y = 1 \mid \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0, & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

Linear Separability

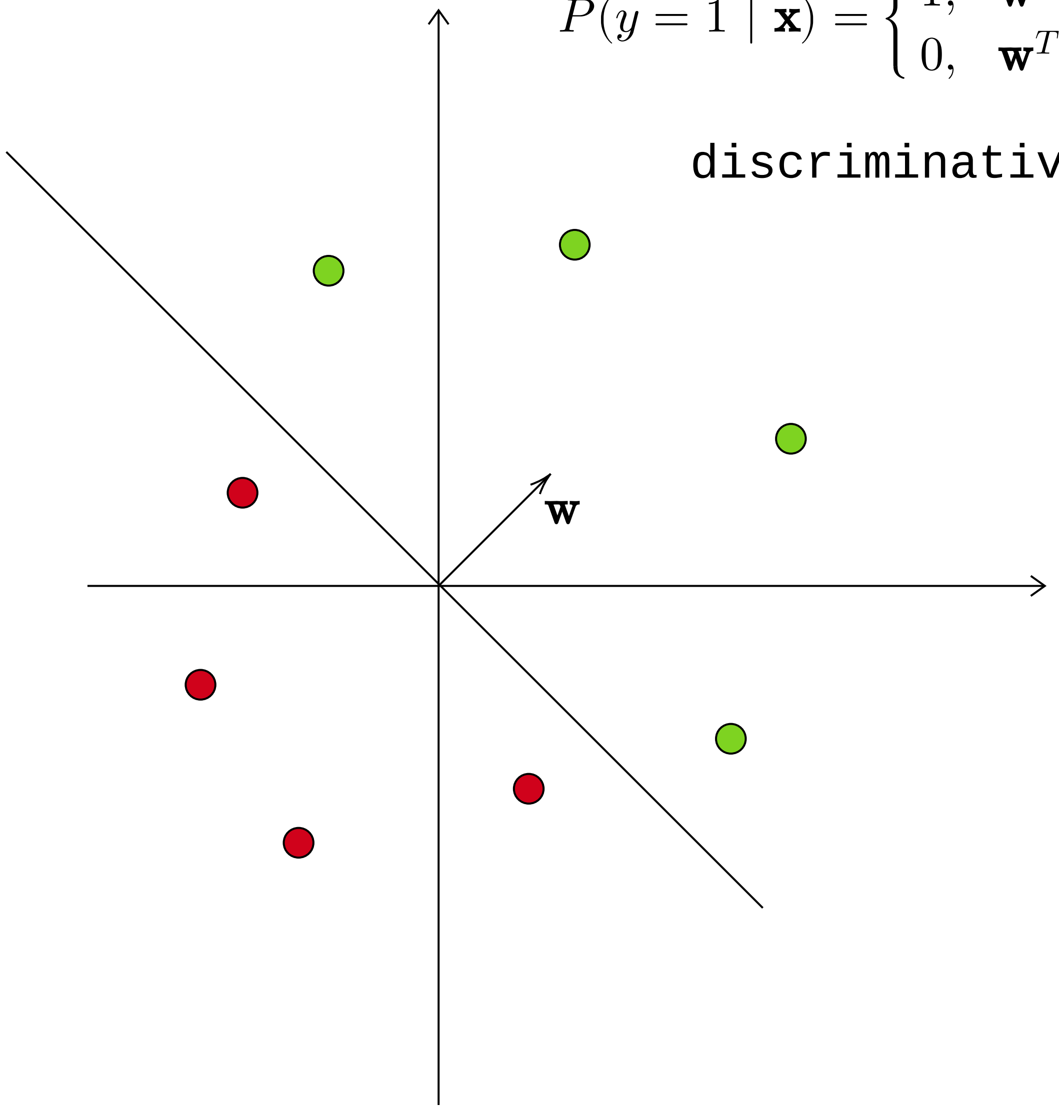
$$P(y = 1 \mid \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0, & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

discriminative

Linear Separability

$$P(y=1 \mid \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0, & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

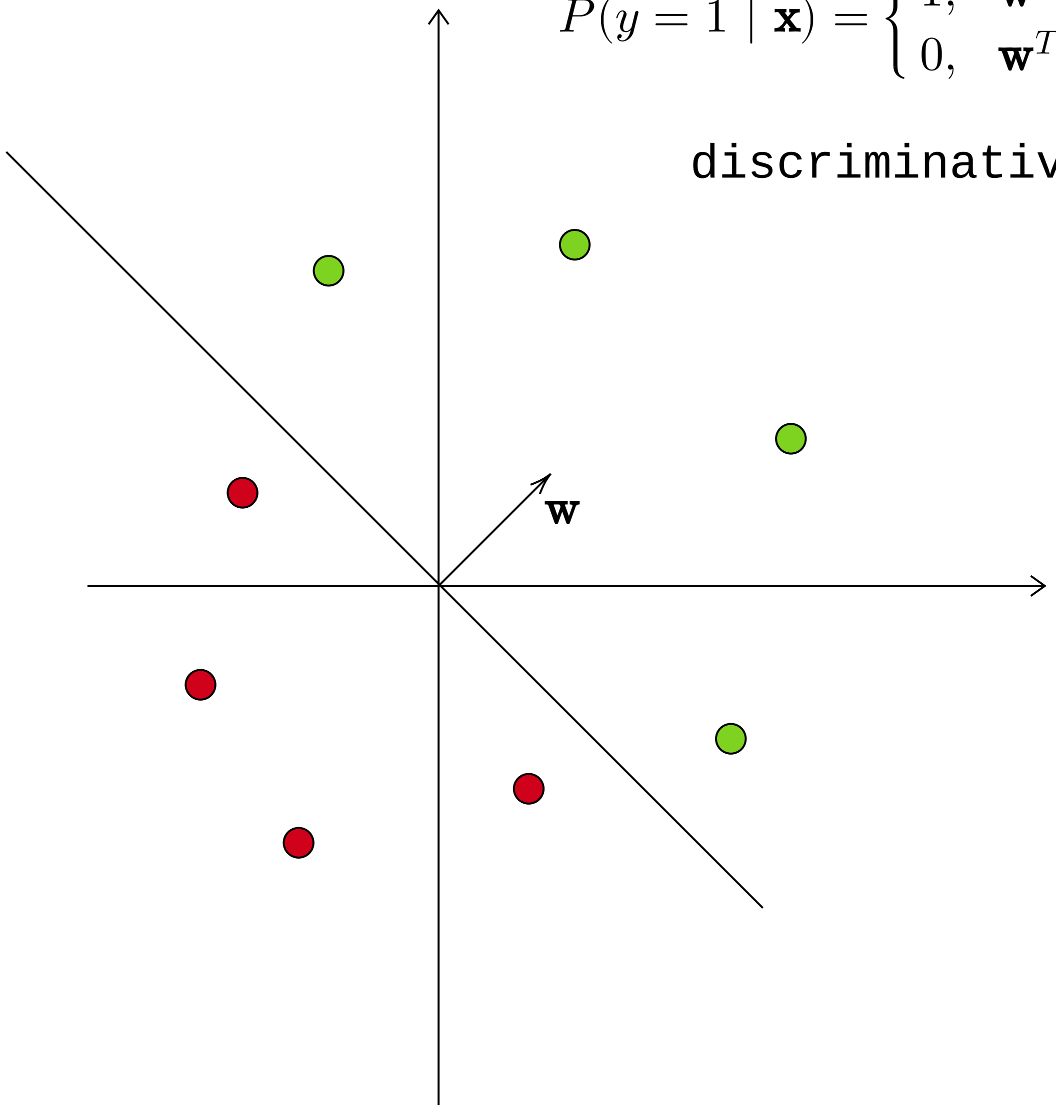
discriminative

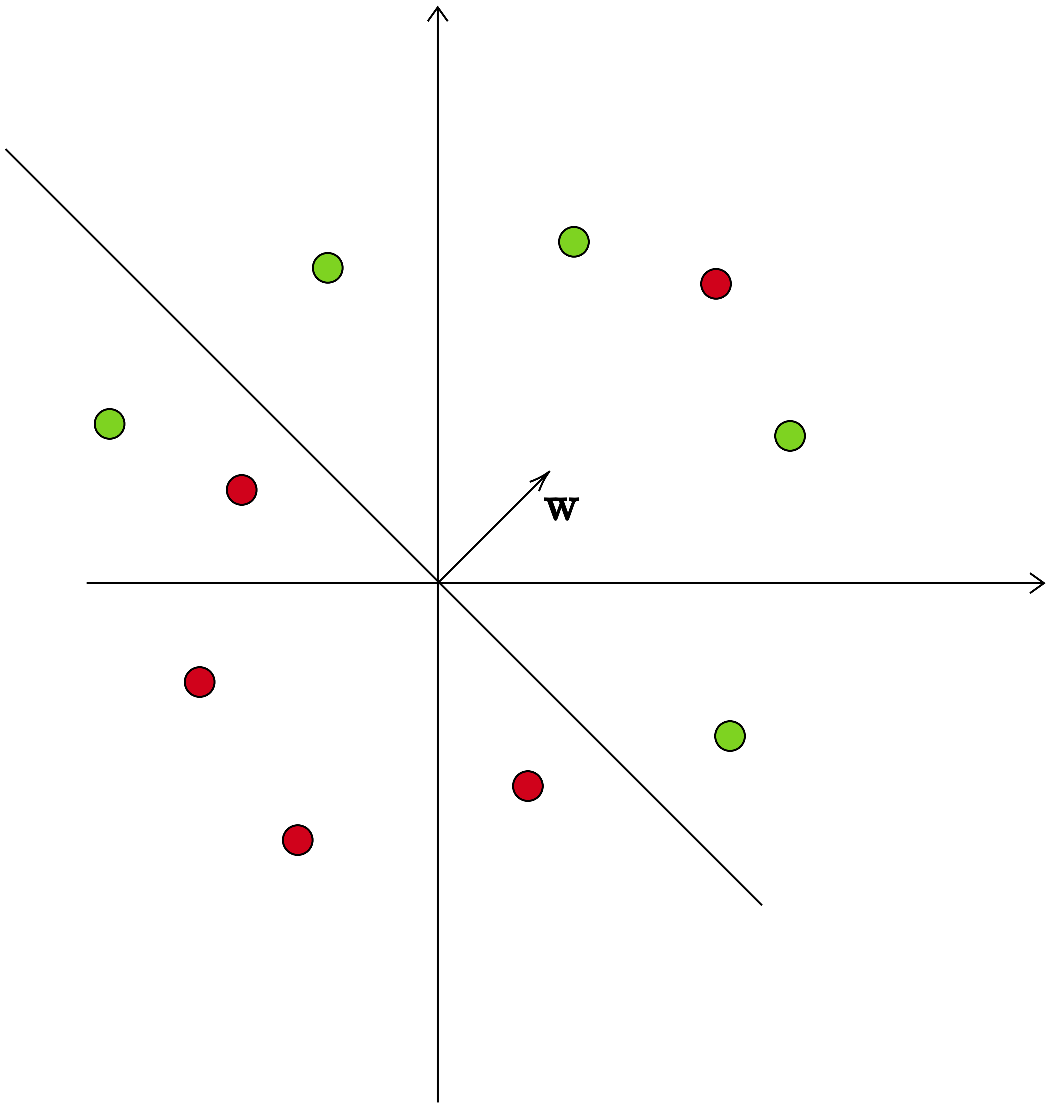


Linear Separability

$$P(y = 1 \mid \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0, & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

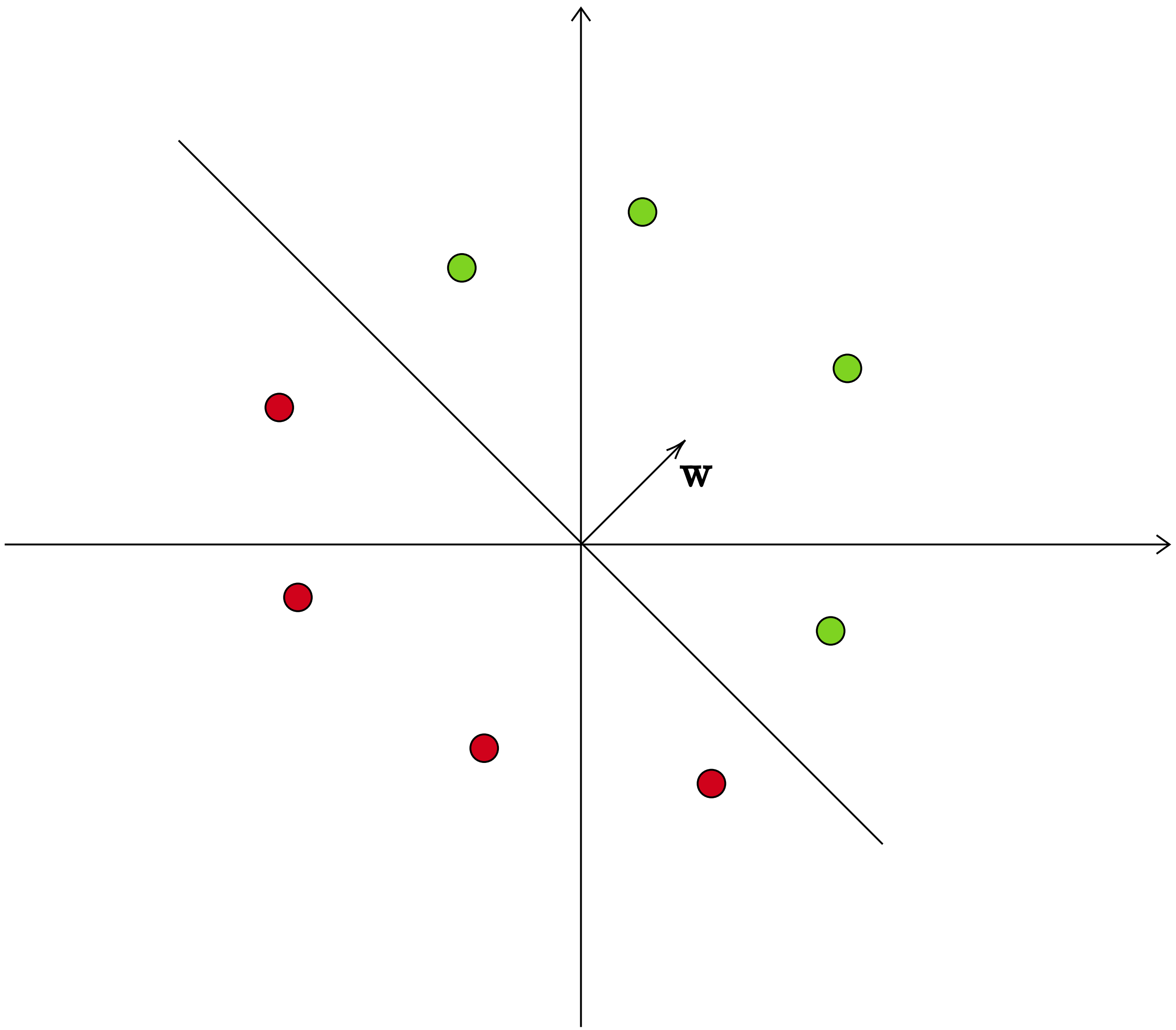
discriminative





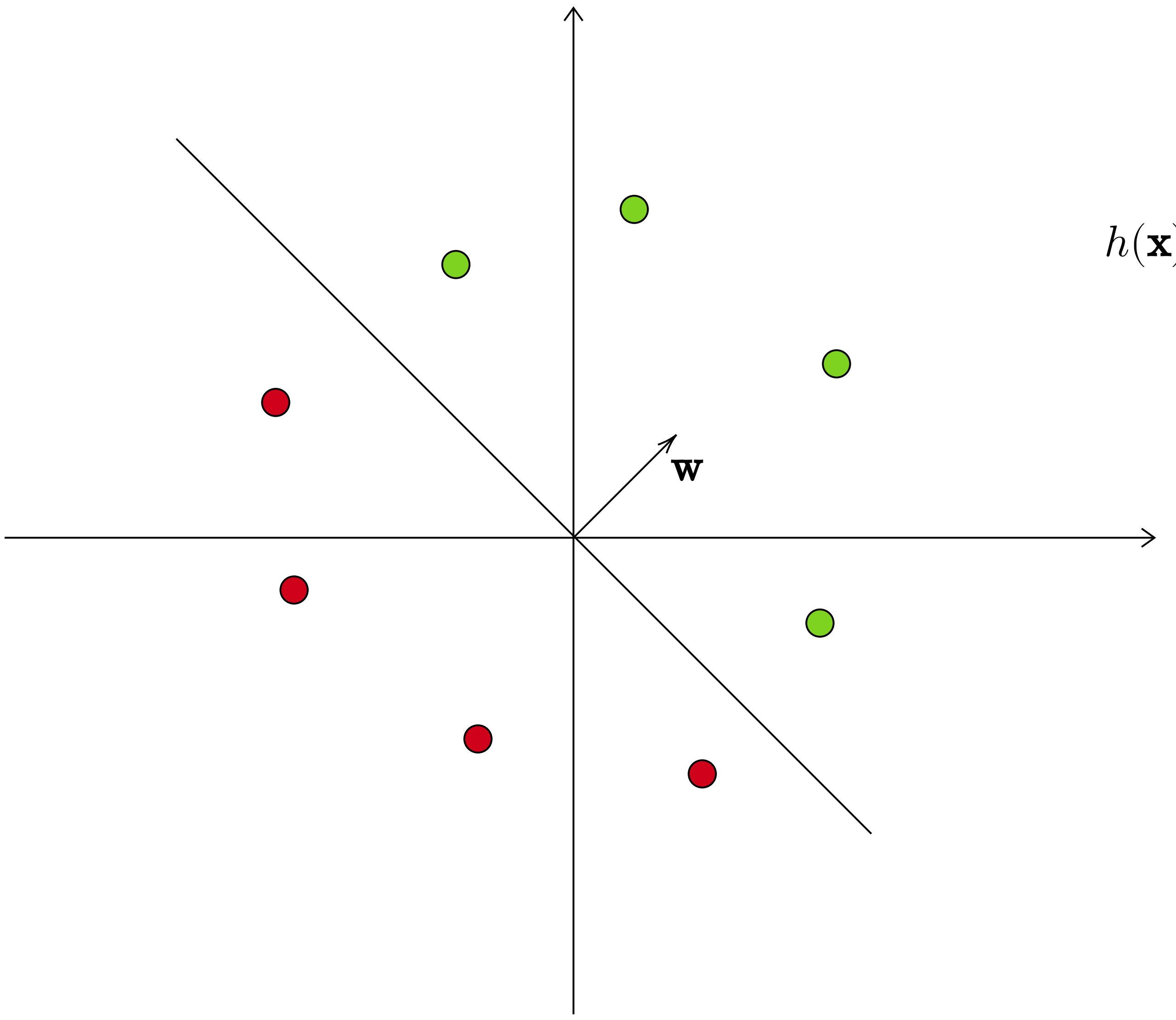
Linear Classifiers (linear separability)

$$D = \{(\mathbf{x}_1, y_1), \ \cdots, (\mathbf{x}_n, y_n)\}$$



Linear Classifiers (linear separability)

$$D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$$



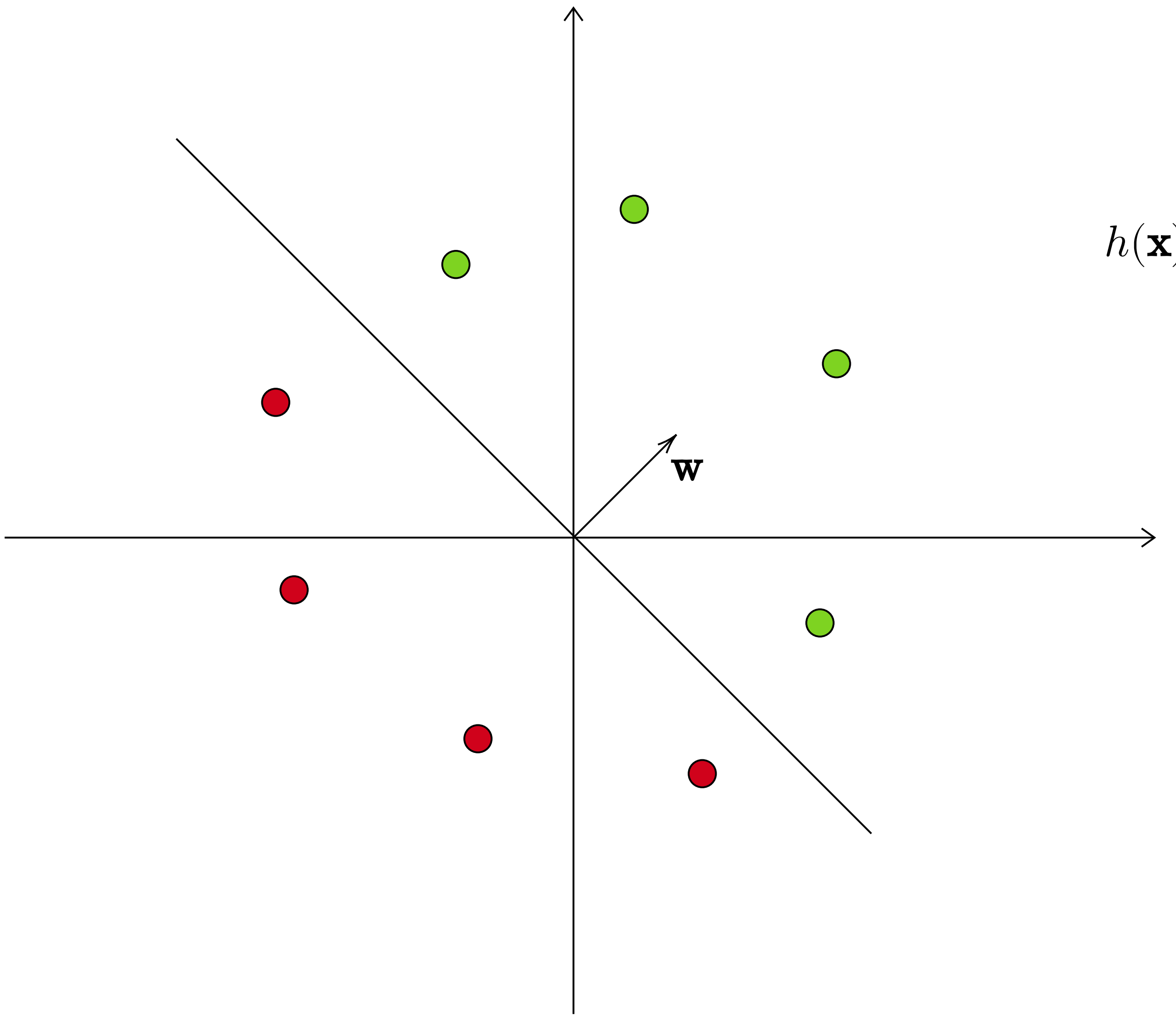
$$h(\mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0 & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

OR

$$h(\mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

Linear Classifiers (linear separability)

$$D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$$



$$h(\mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ 0 & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

OR

$$h(\mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

$$\min_{h \in \mathcal{H}_{\texttt{linear}}} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{1}[h(\mathbf{x}_i) \neq y_i]$$

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$(1)\ \ \texttt{Perceptron}(D):$$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

(1) $\texttt{Perceptron}(D):$

(2) $\quad \mathbf{w}^0 = \mathbf{0}$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t\,T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t\,T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t\,T}\mathbf{x} < 0 \end{cases}$$

(1) $\texttt{Perceptron}(D):$

(2) $\quad\mathbf{w}^0 = \mathbf{0}$

(3) `classified = False`

Perceptron

$$D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

(1) Perceptron(D):

(2) $\quad \mathbf{w}^0 = \mathbf{0}$

(3) $\quad$ classified = False

(4) $\quad$ while not classified:

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\mathtt{sign}\left(\mathbf{w}^{t\,T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t\,T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t\,T}\mathbf{x} < 0 \end{cases}$$

(1) `Perceptron`(D):

(2) $\quad \mathbf{w}^0 = \mathbf{0}$

(3) `    classified = False`

(4) `    while not classified:`

(5) `        classified = True`

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t\,T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t\,T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t\,T}\mathbf{x} < 0 \end{cases}$$

```
Perceptron(D):
    w^0 = 0
    classified = False
    while not classified:
        classified = True
        for i = 1 to i = n:
```

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\mathtt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

$\mathtt{Perceptron}(D):$

    $\mathbf{w}^0 = \mathbf{0}$

    `classified = False`

    `while not classified:`

        `classified = True`

        `for` $i = 1$ `to` $i = n:$

            `if` $\mathtt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}_i\right) \neq y_i:$

Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

$\texttt{Perceptron}(D):$

    $\mathbf{w}^0 = \mathbf{0}$

    `classified = False`

    `while not classified:`

        `classified = True`

        `for` $i = 1$ `to` $i = n:$

            `if` $\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}_i\right) \neq y_i:$

                `classified = False`

Perceptron

$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$

$$\text{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

Perceptron(D):

  $\mathbf{w}^0 = \mathbf{0}$

  classified = False

  while not classified:

    classified = True

    for $i = 1$ to $i = n$:

      if sign$\left(\mathbf{w}^{t^T}\mathbf{x}_i\right) \neq y_i$:

        classified = False

        $\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}_i y_i$

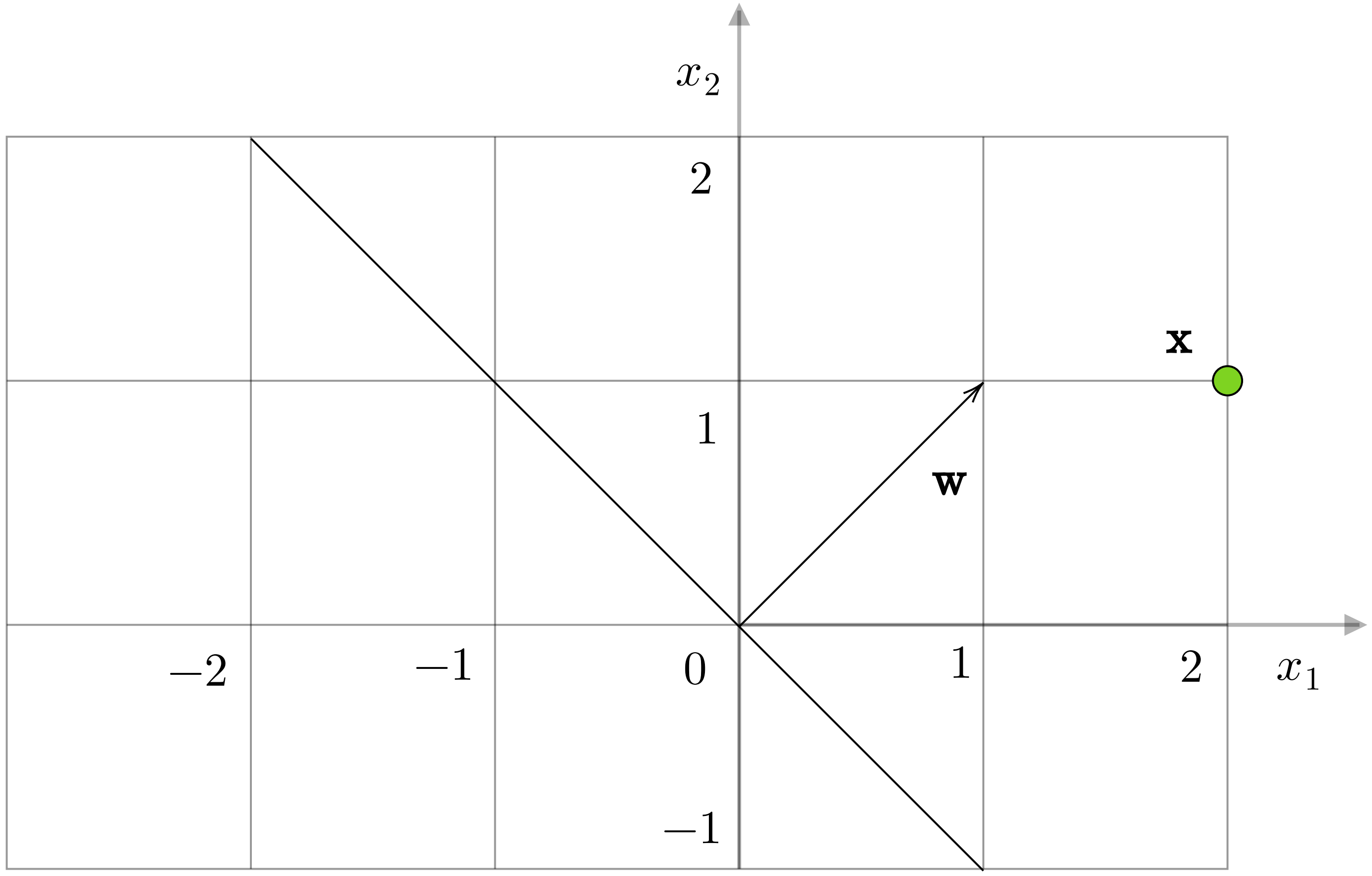
Perceptron

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\}$$

$$\texttt{sign}\left(\mathbf{w}^{t^T}\mathbf{x}\right) = \begin{cases} 1, & \mathbf{w}^{t^T}\mathbf{x} \geqslant 0 \\ -1 & \mathbf{w}^{t^T}\mathbf{x} < 0 \end{cases}$$

`Perceptron`(D) :
    $\mathbf{w}^0 = \mathbf{0}$
    `classified = False`
    `while not classified:`
        `classified = True`
        `for` $i = 1$ `to` $i = n$:
            `if sign`$\left(\mathbf{w}^{t^T}\mathbf{x}_i\right) \neq y_i$:
                `classified = False`
                $\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}_i y_i$
    `return` $\mathbf{w}$

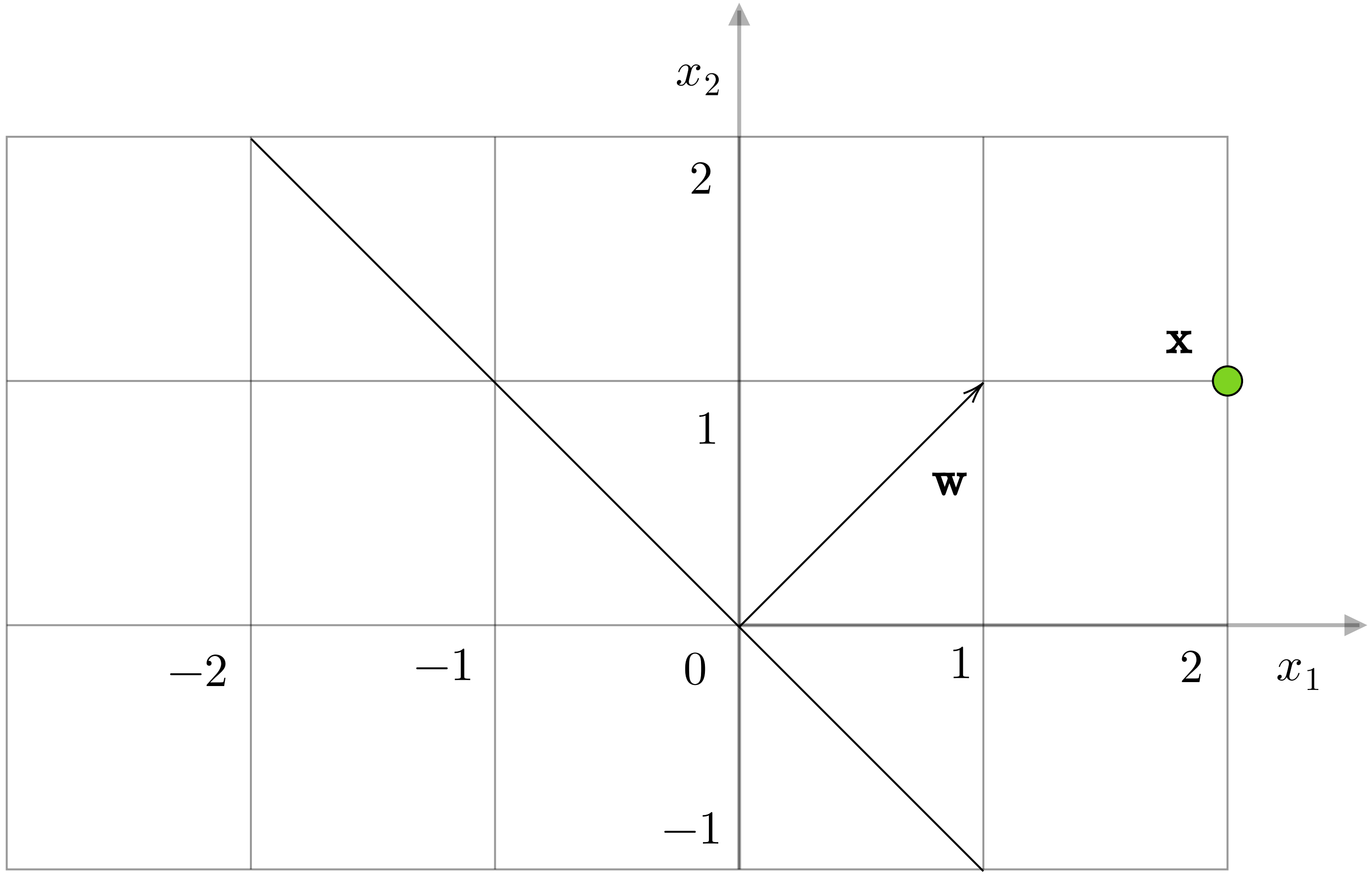
Classifiying Outcomes



Actual =

Predicted =

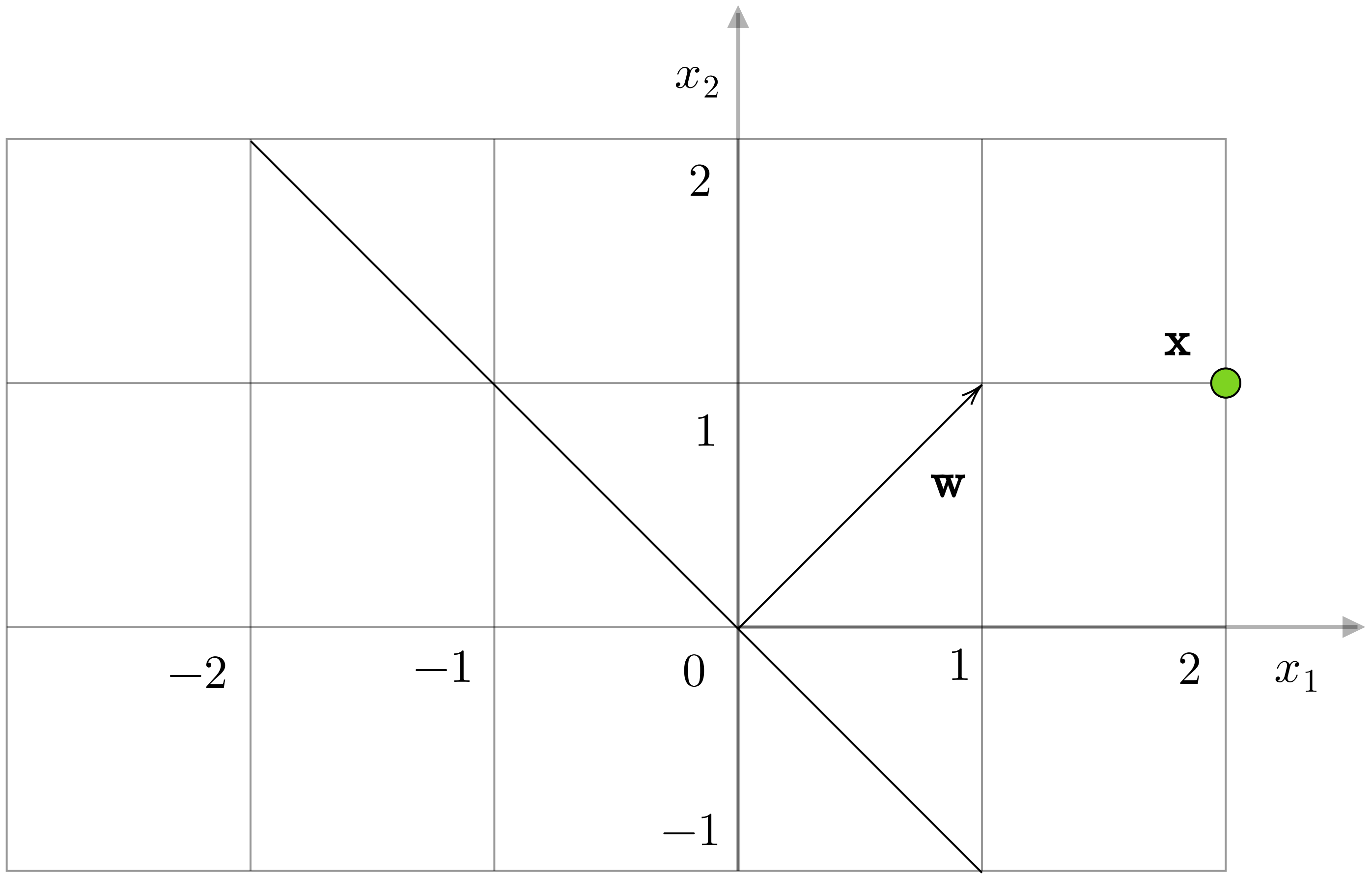
Classifiying Outcomes



Actual = 1

Predicted = 1

Classifiying Outcomes

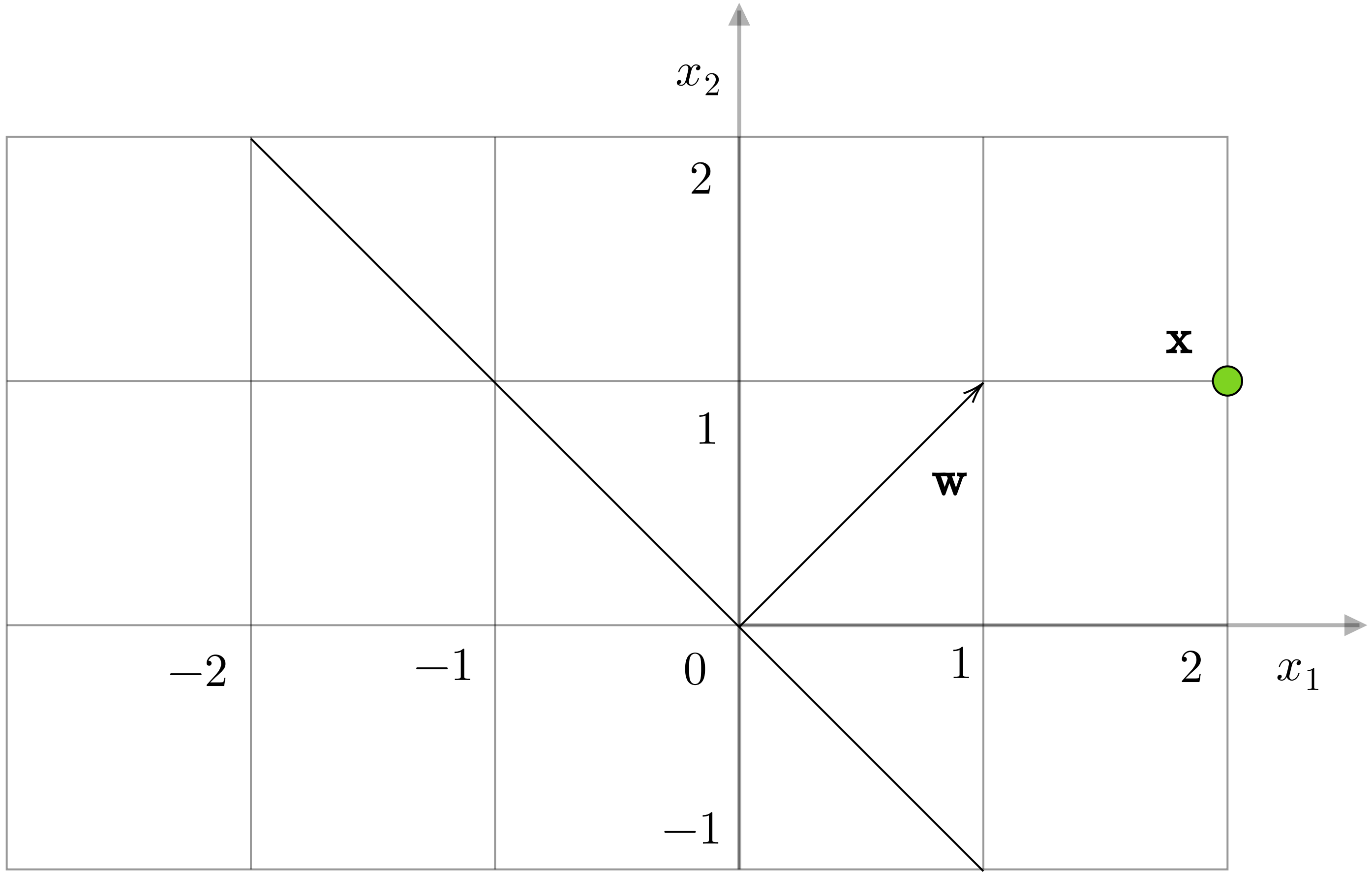


Actual = 1

Predicted = 1

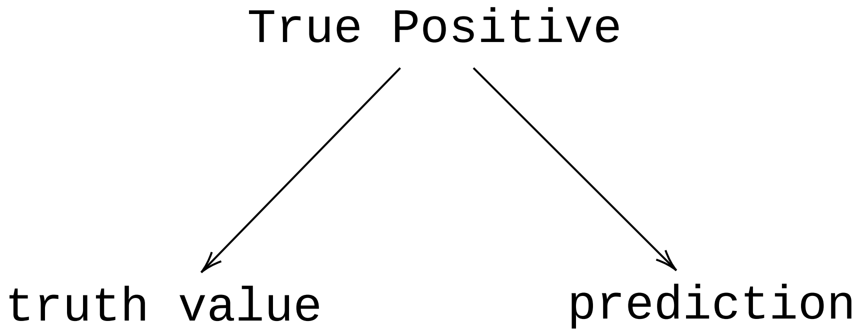
True Positive

Classifiying Outcomes

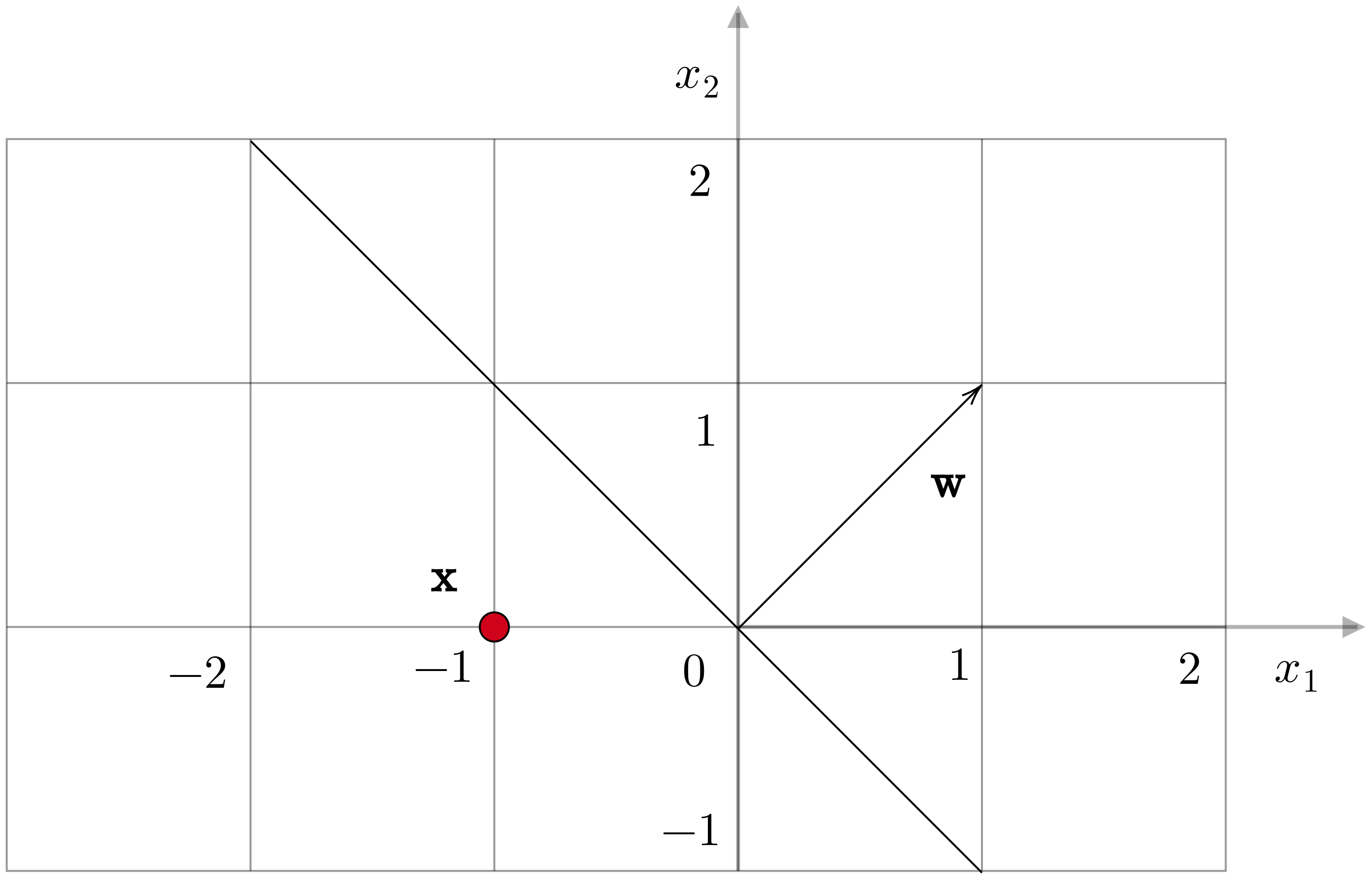


Actual = 1

Predicted = 1



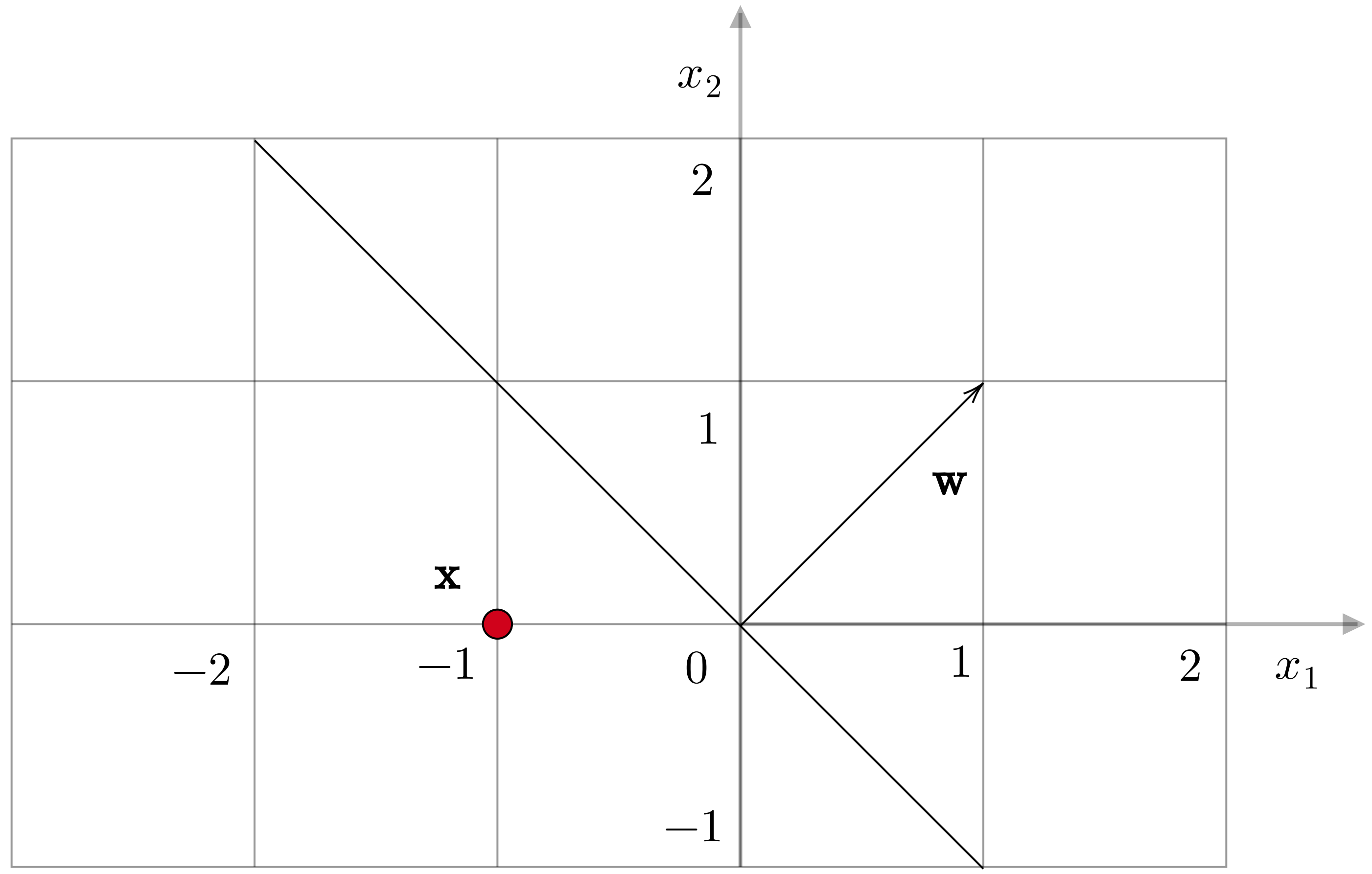
Classifiying Outcomes



Actual =

Predicted =

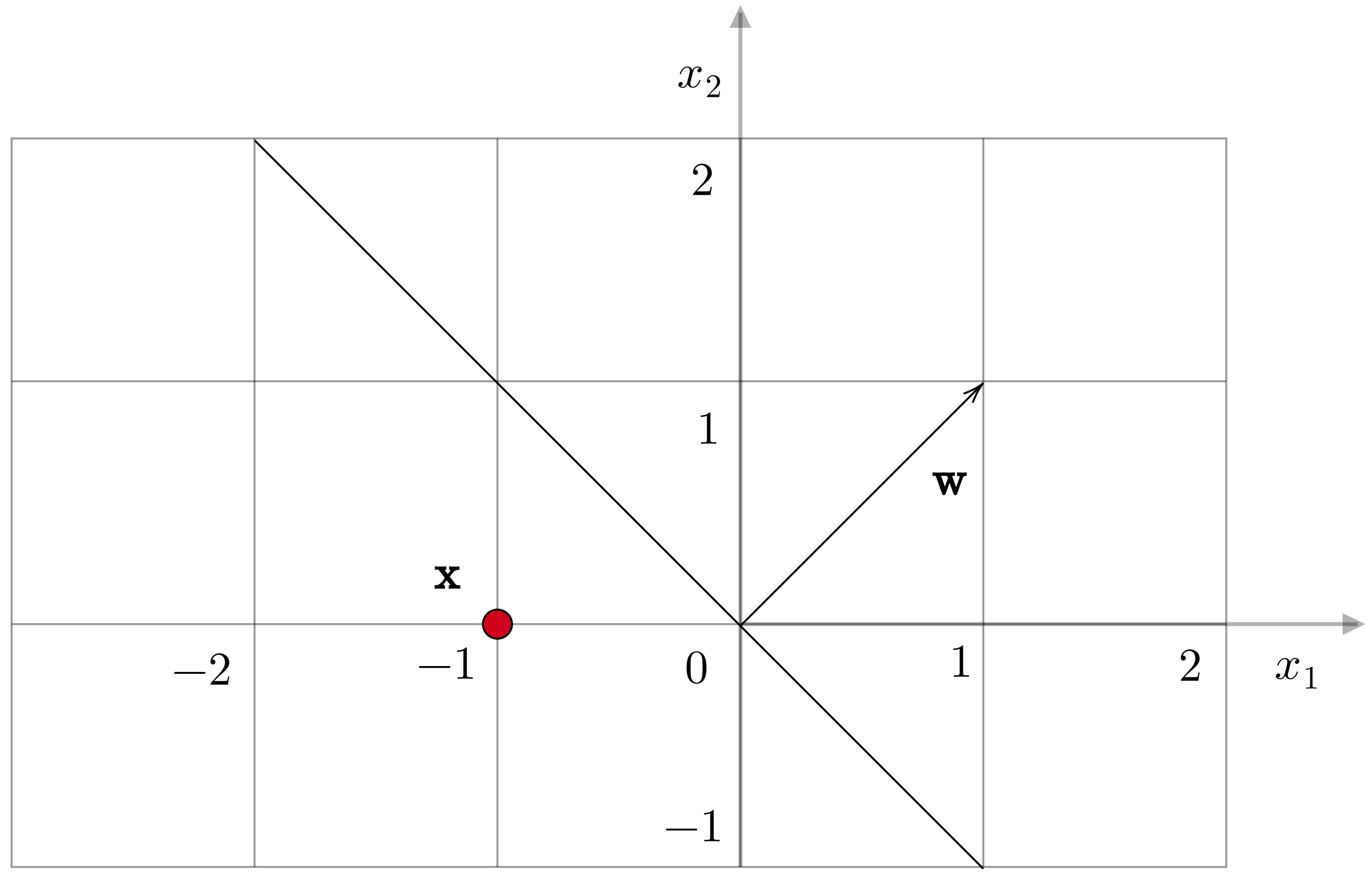
Classifiying Outcomes



Actual = -1

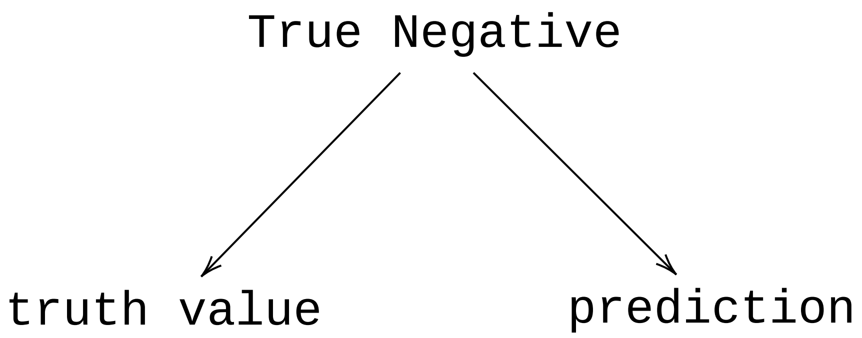
Predicted = -1

Classifiying Outcomes

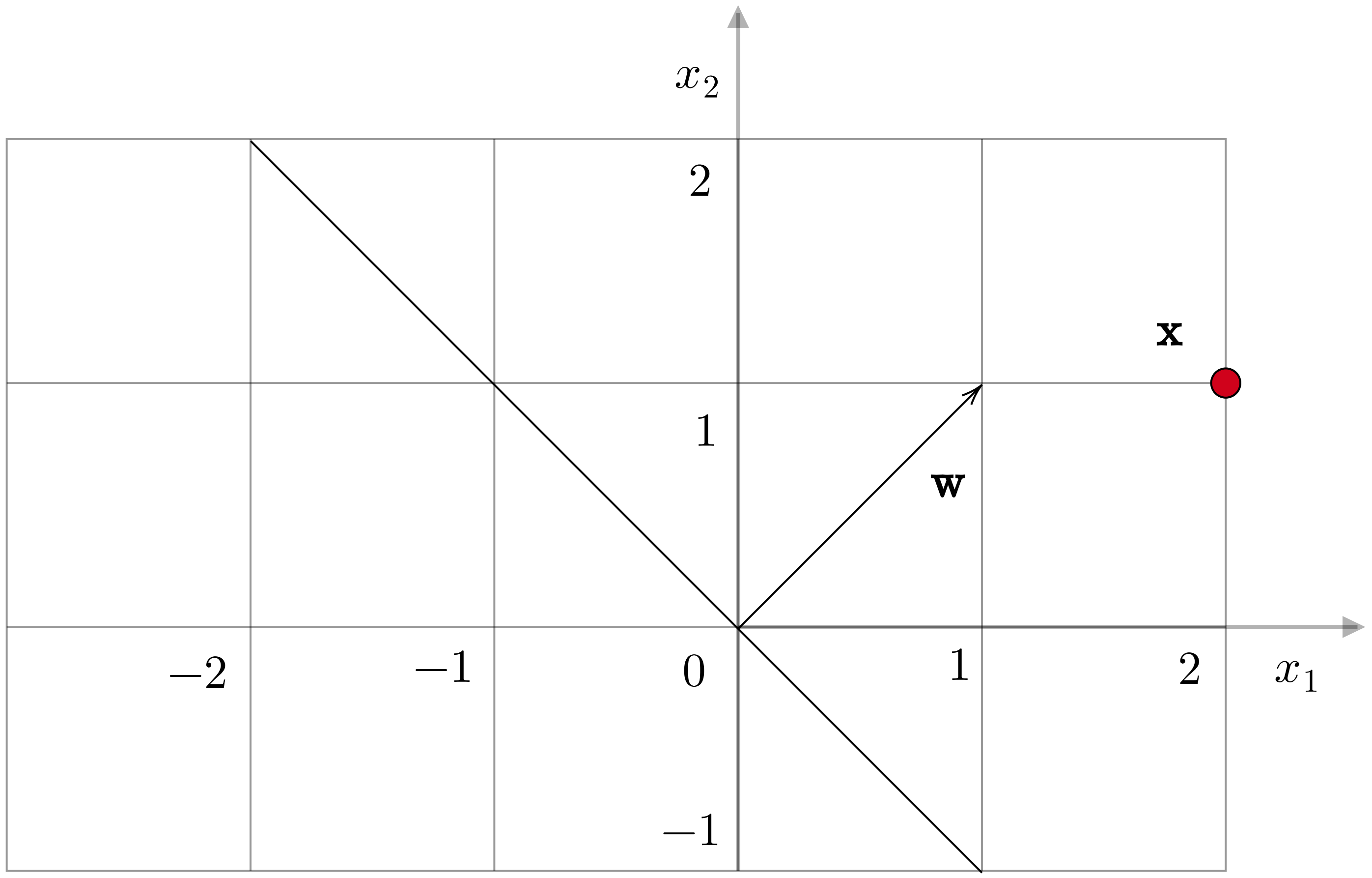


Actual = -1

Predicted = -1



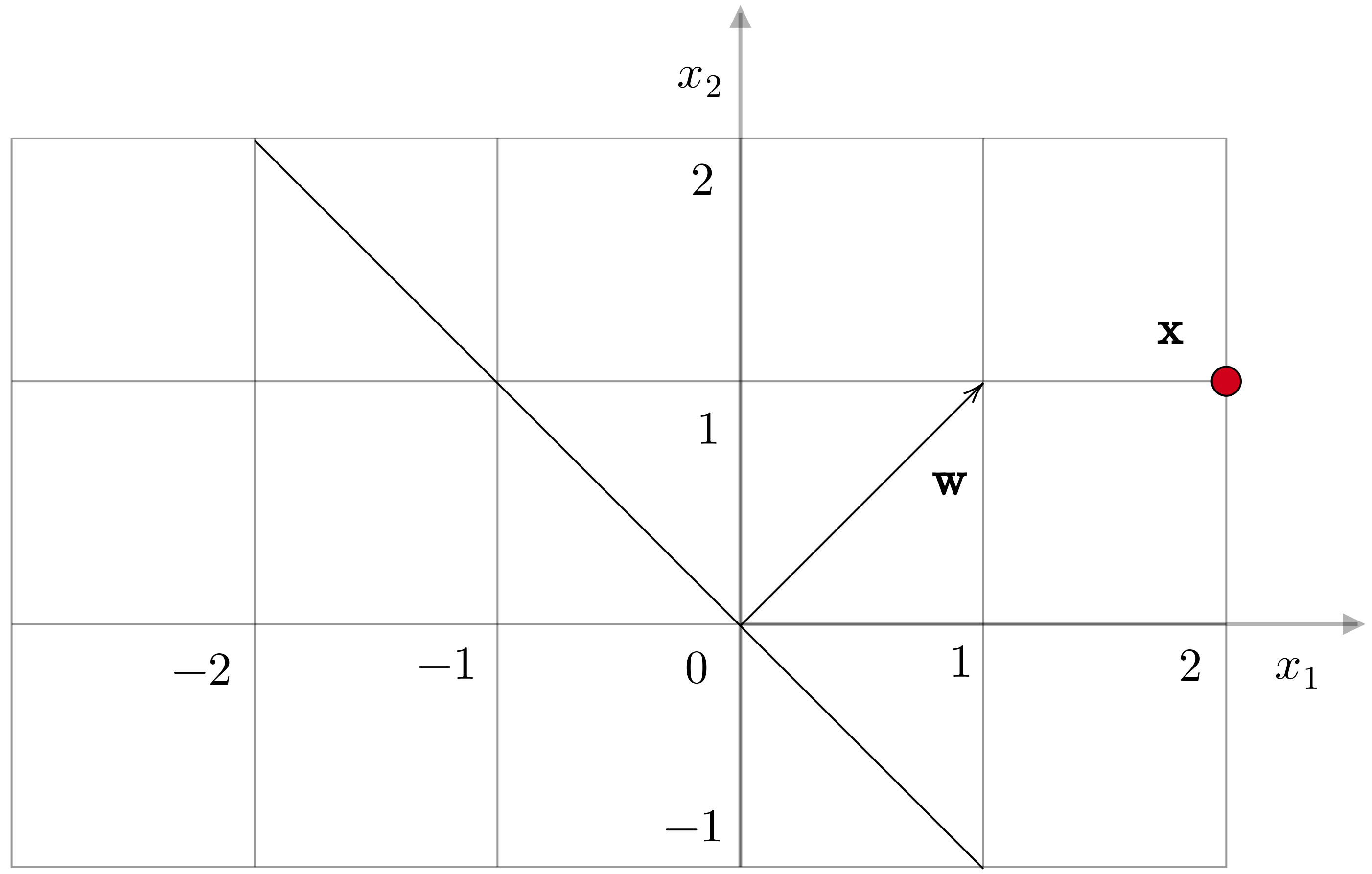
Classifiying Outcomes



Actual =

Predicted =

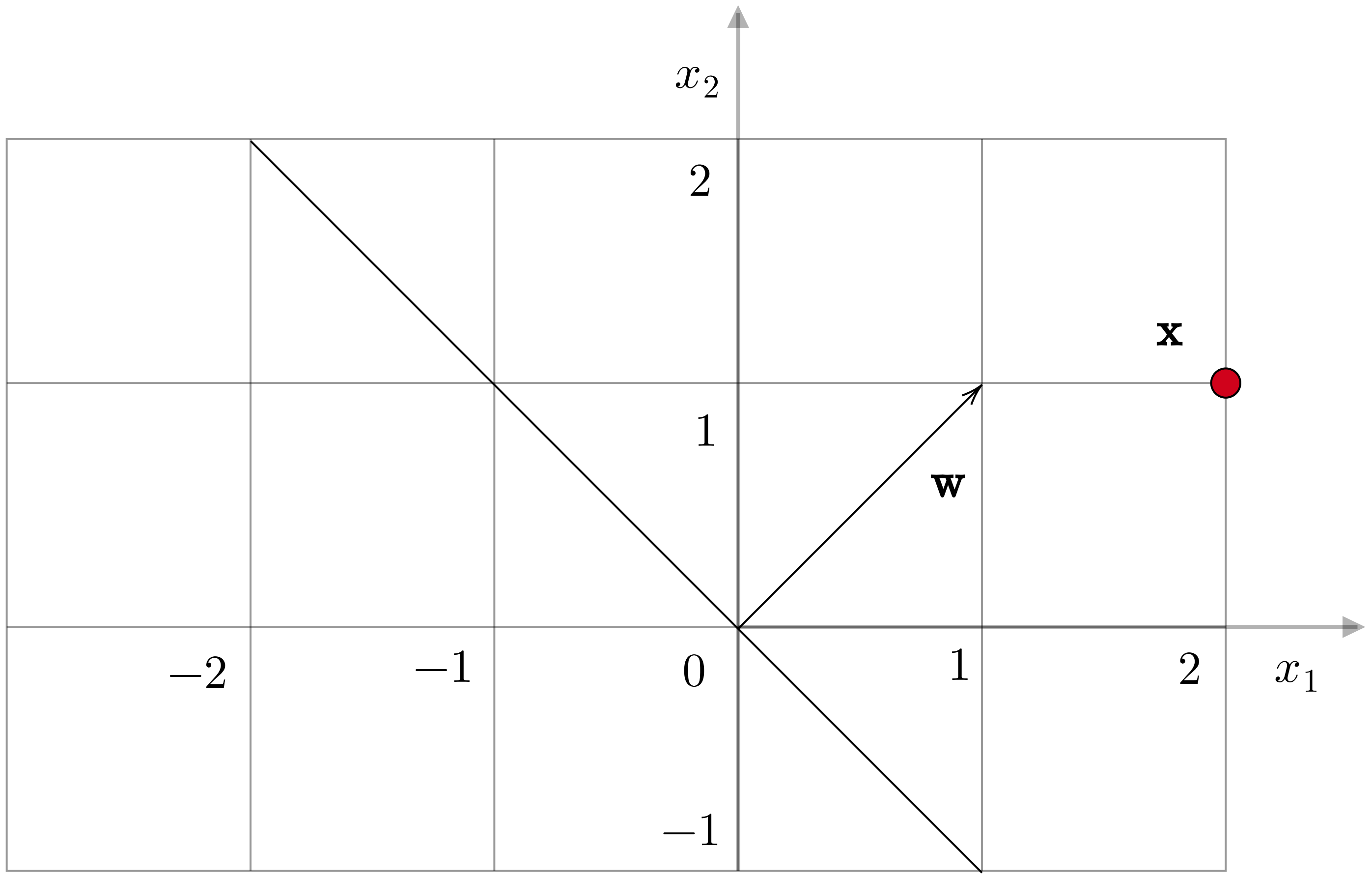
Classifiying Outcomes



Actual = -1

Predicted = 1

Classifiying Outcomes



Actual = -1

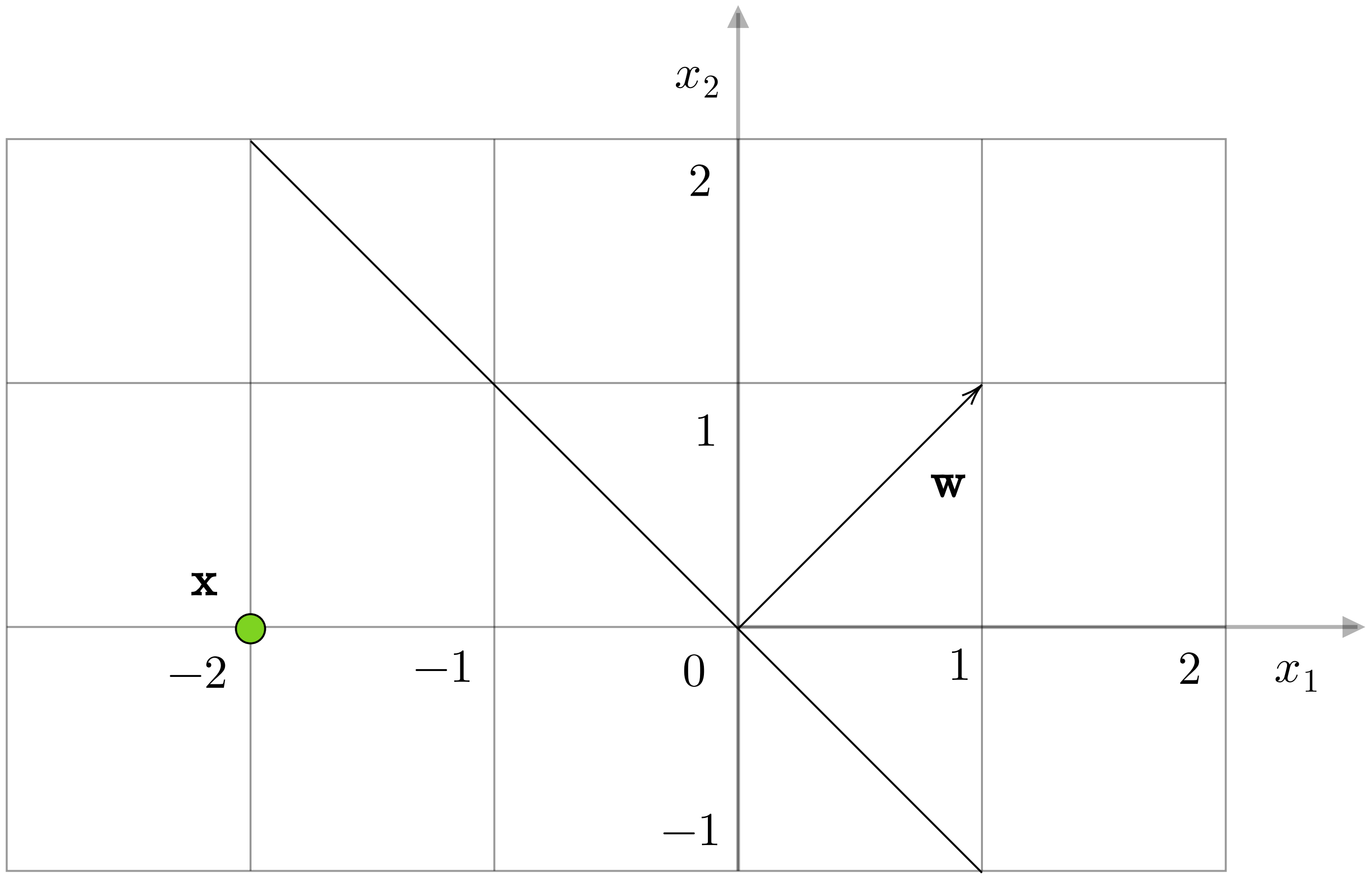
Predicted = 1

False Positive

truth value

prediction

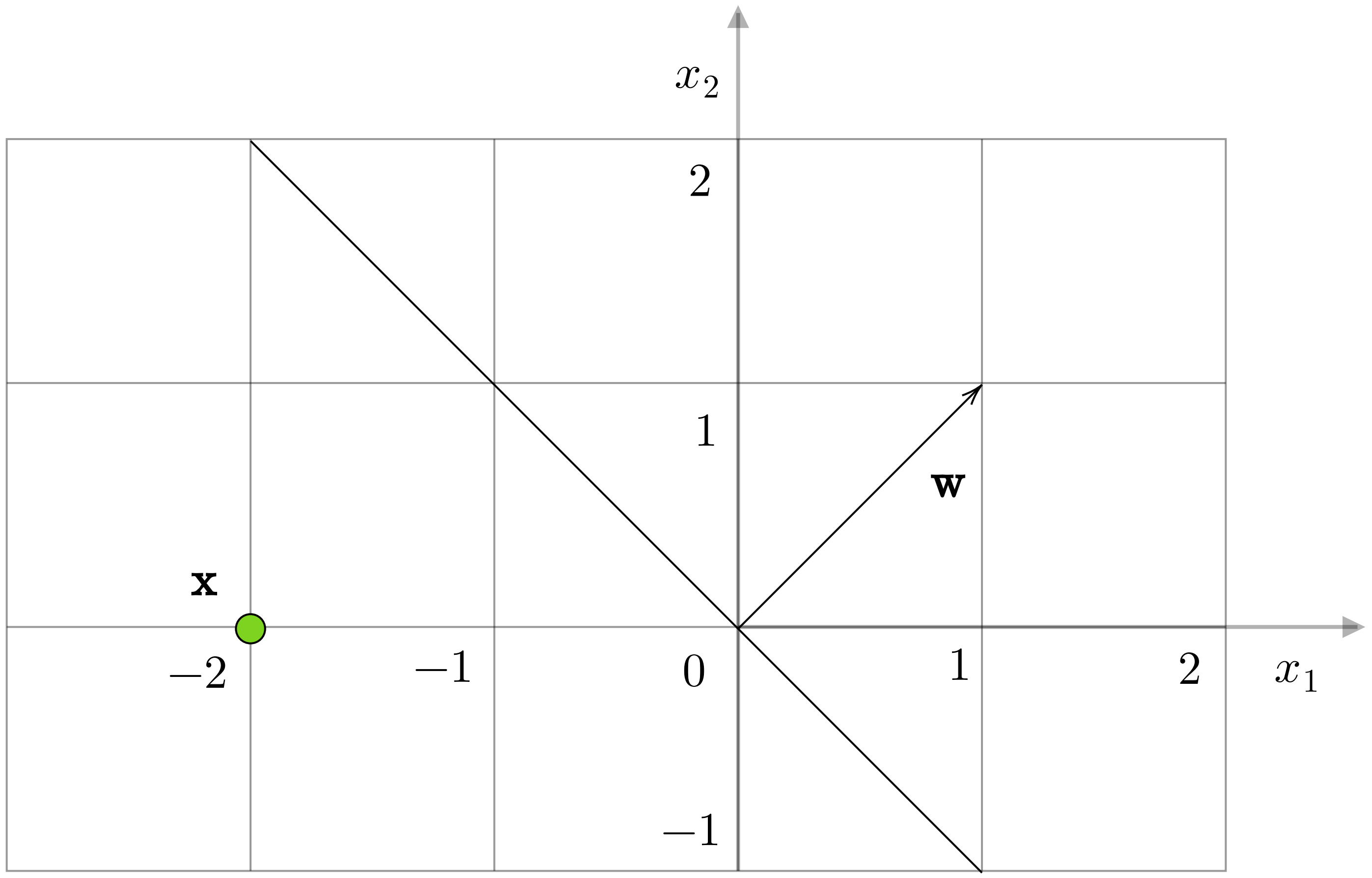
Classifiying Outcomes



Actual =

Predicted =

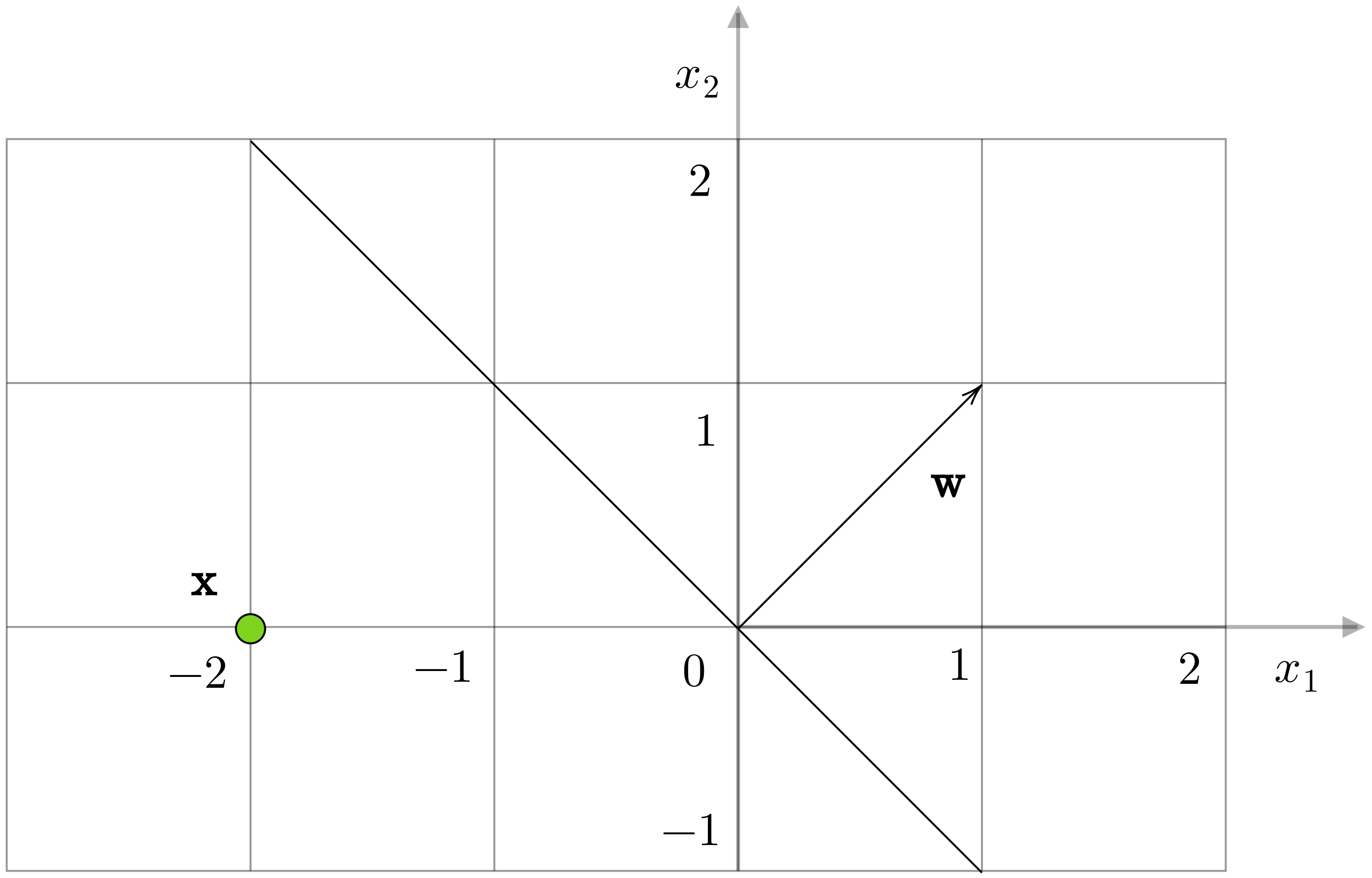
Classifiying Outcomes



Actual = 1

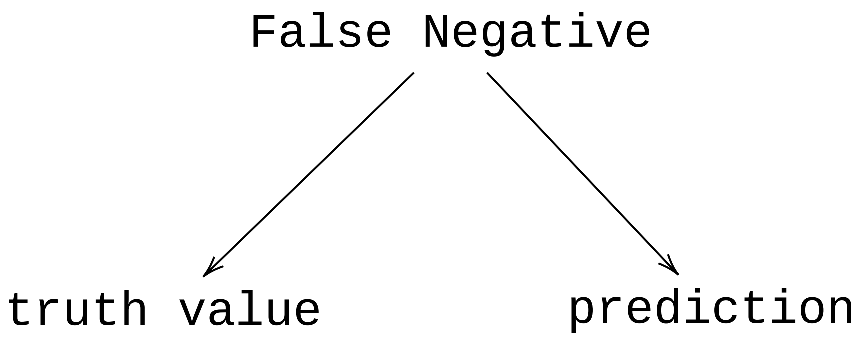
Predicted = -1

Classifiying Outcomes



Actual = 1

Predicted = -1



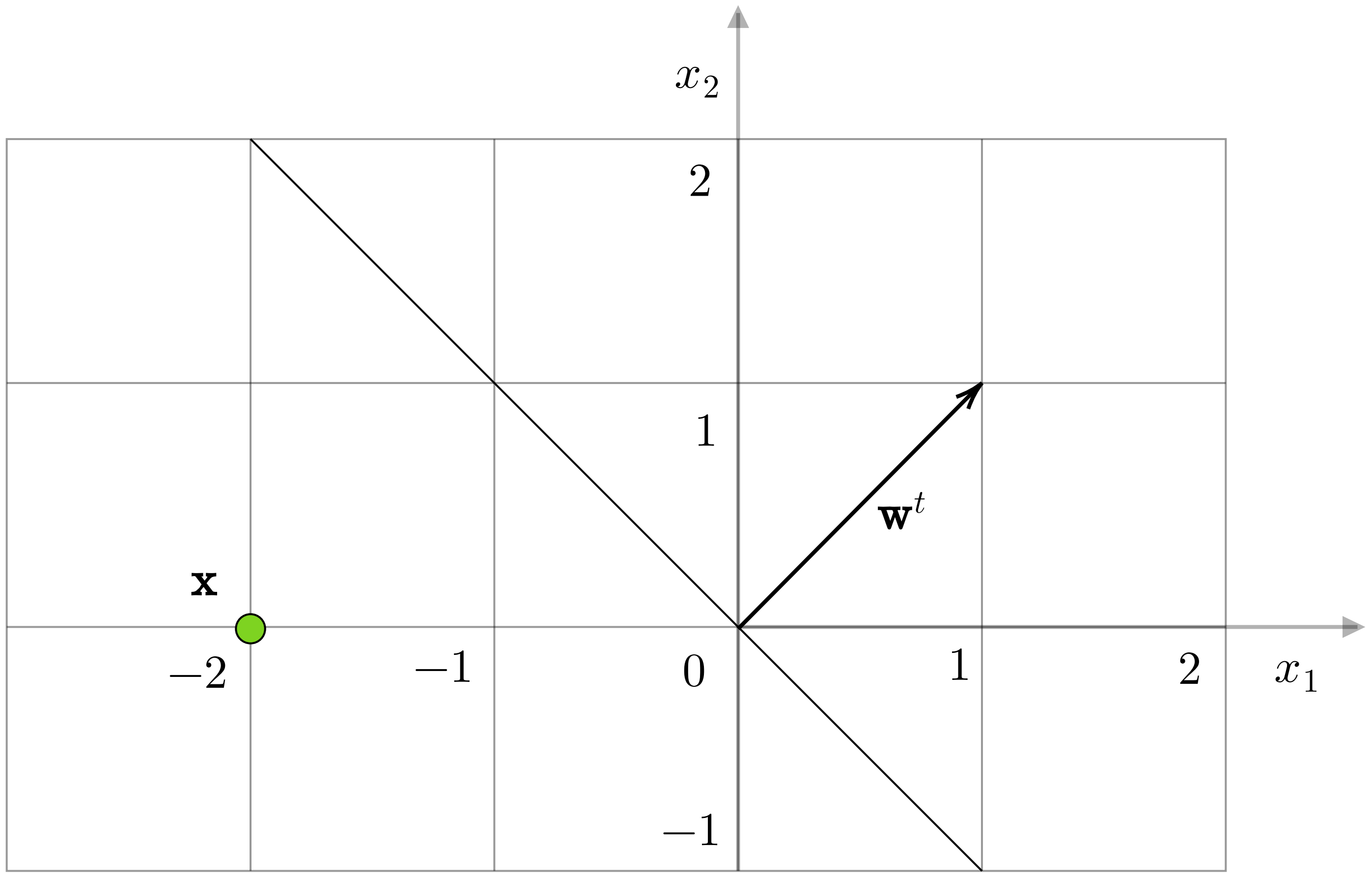
Confusion Matrix

Prediction

Actual

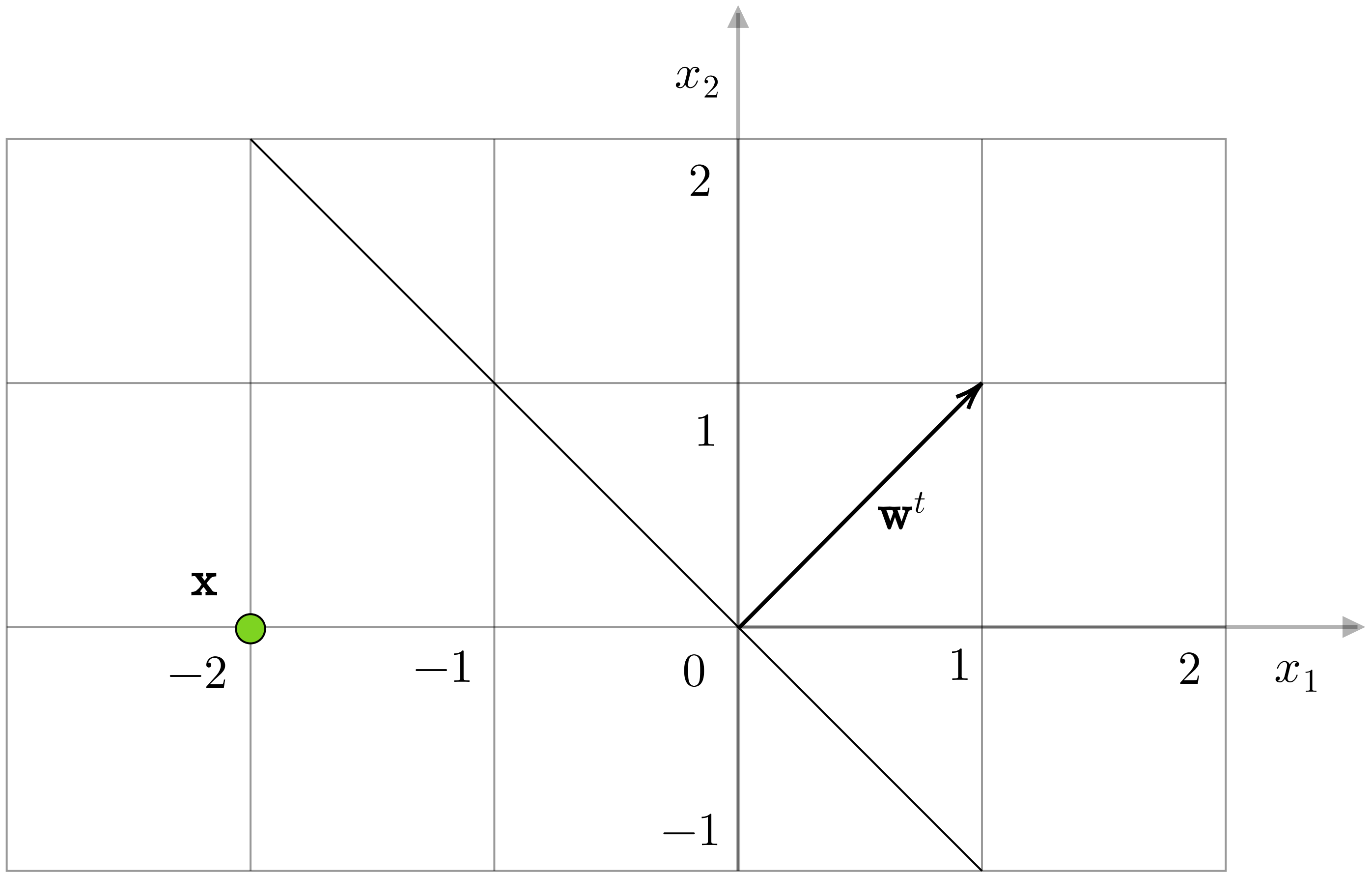
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Perceptron Update Rule



Perceptron Update Rule

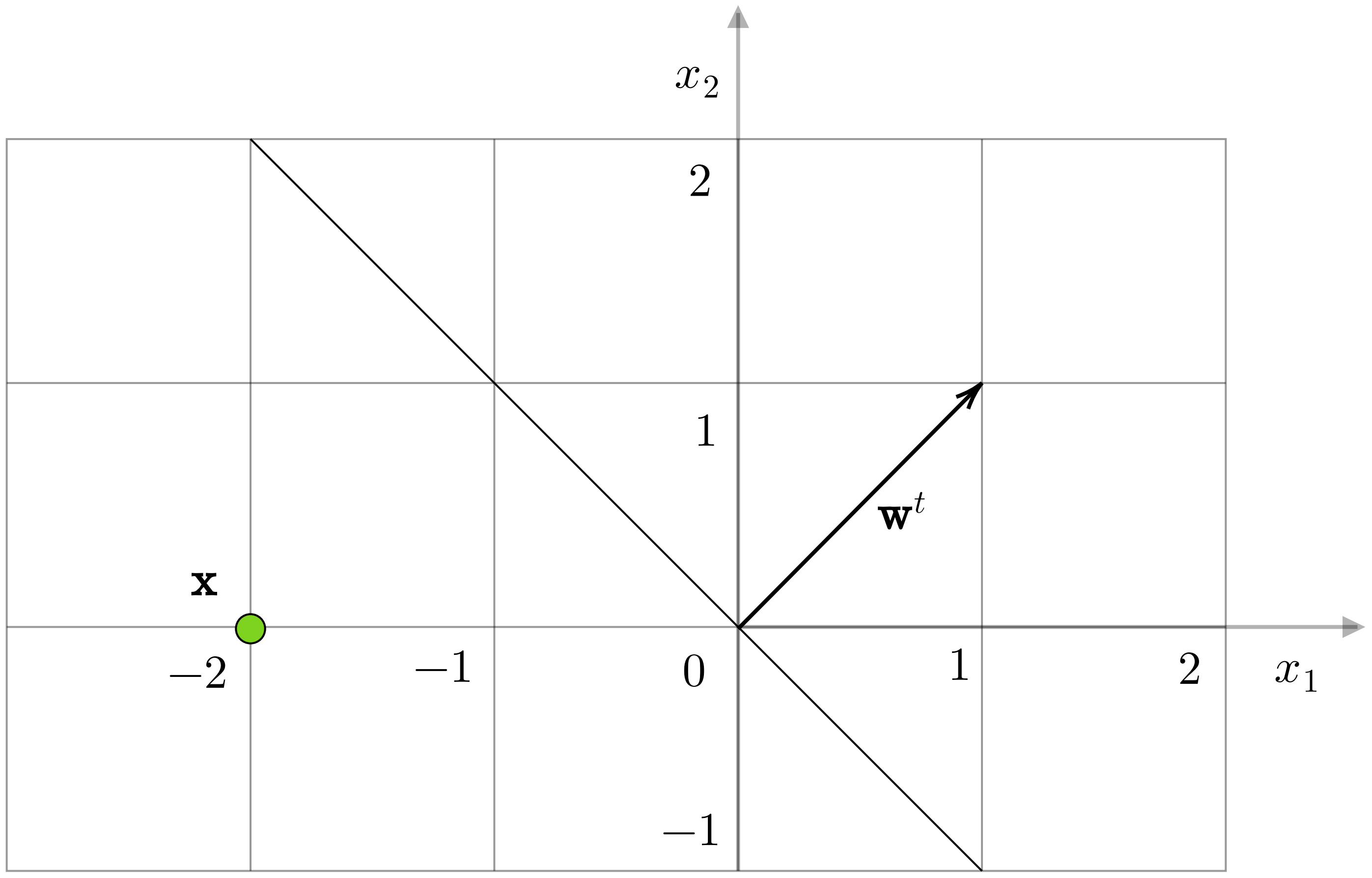
False Negative



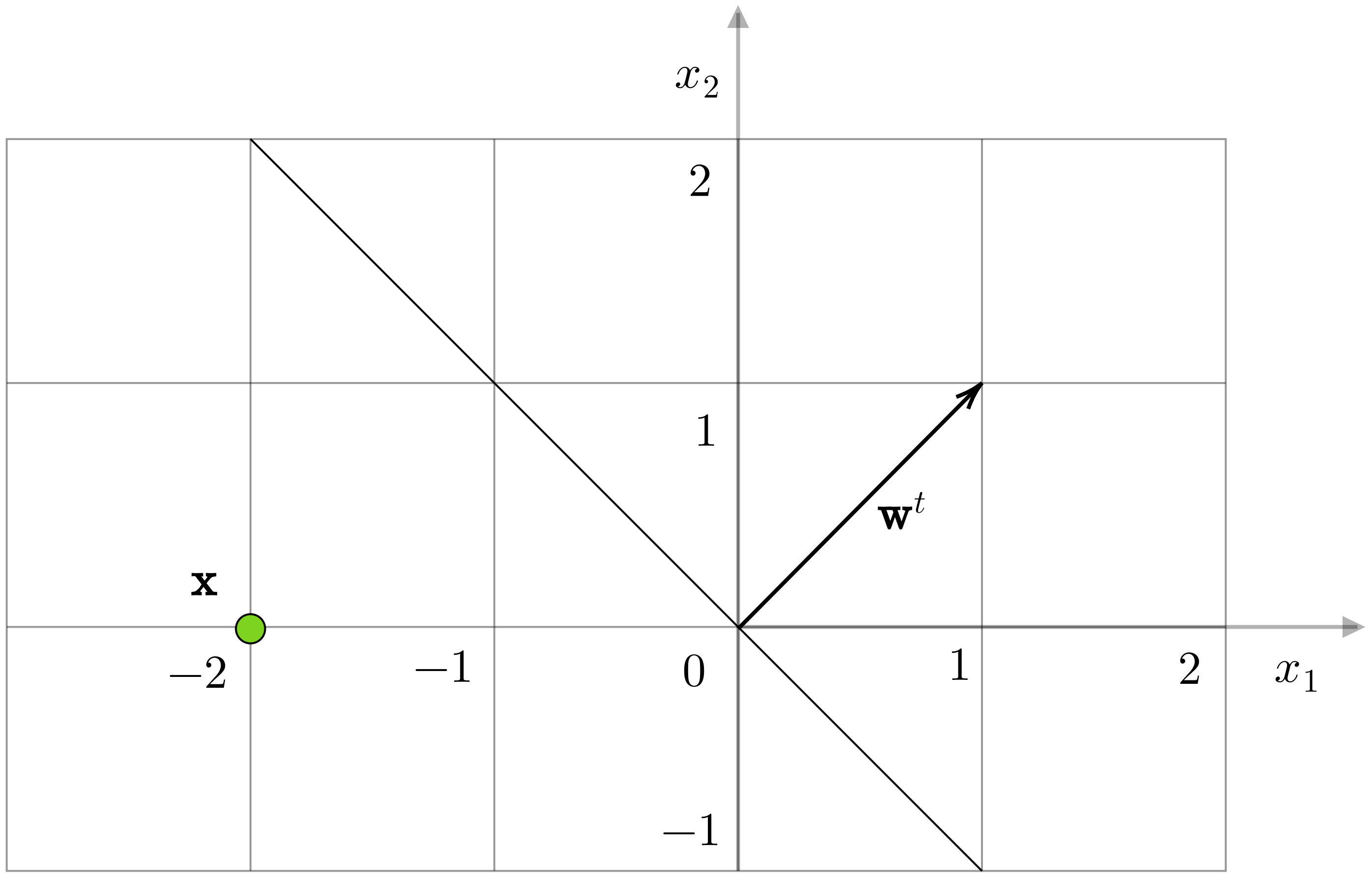
Perceptron Update Rule

False Negative

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$



Perceptron Update Rule



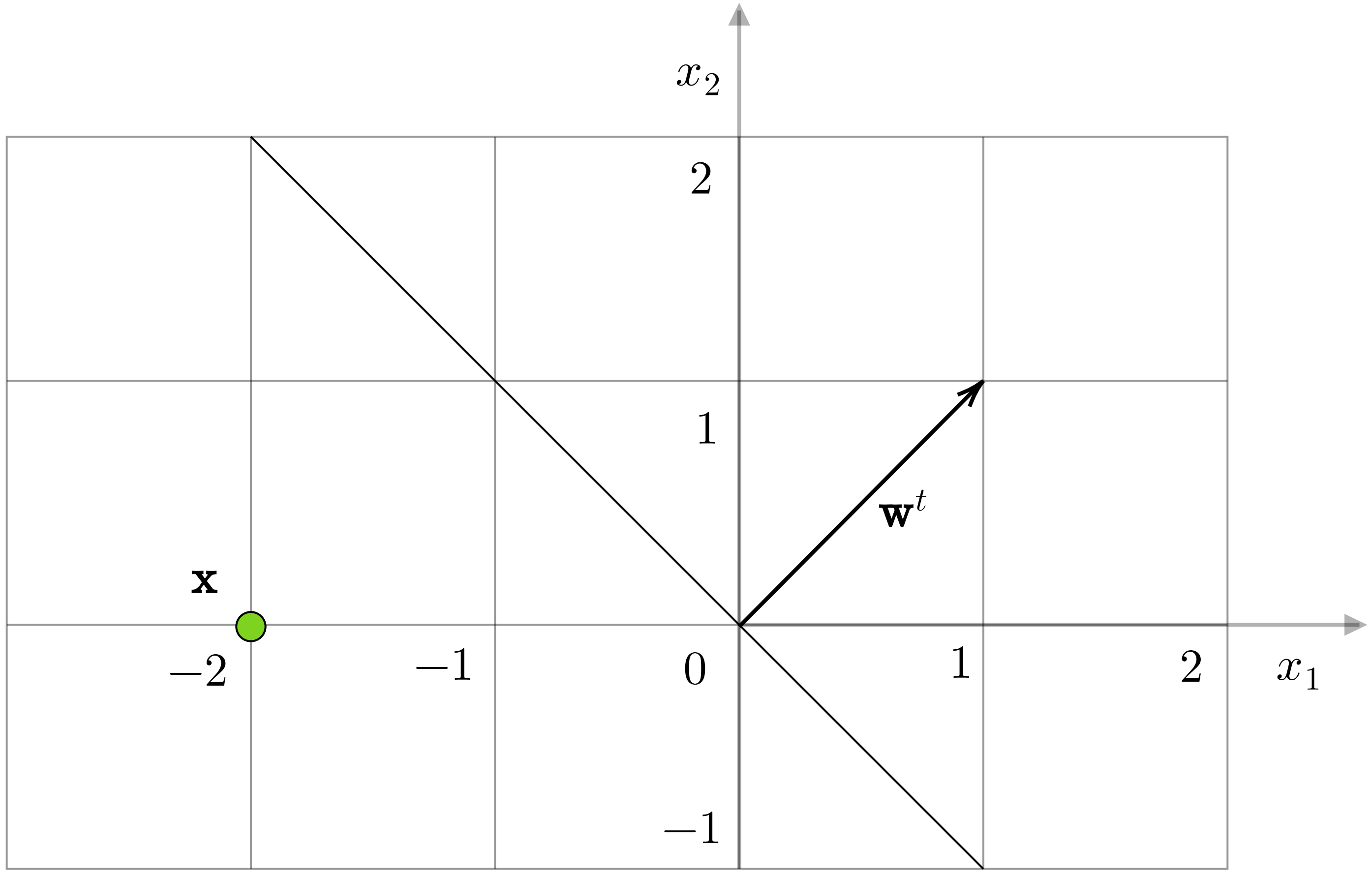
False Negative

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$${\mathbf{w}^t}^T \mathbf{x} = -2$$

Actual = 1 Predicted = -1

Perceptron Update Rule



False Negative

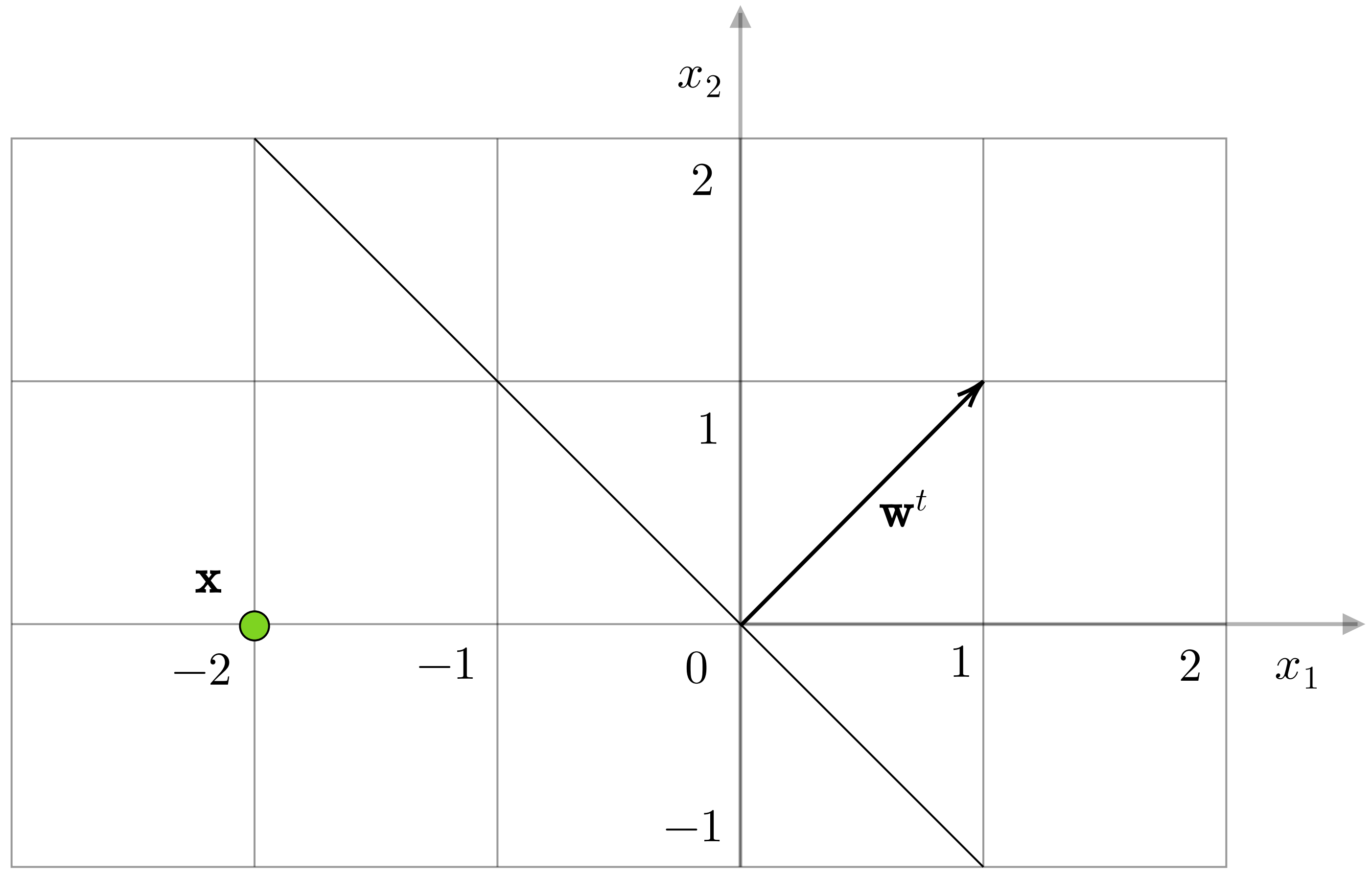
$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$${\mathbf{w}^t}^T \mathbf{x} = -2$$

Actual = 1 Predicted = -1

$$\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}y$$

Perceptron Update Rule



False Negative

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$${\mathbf{w}^t}^T \mathbf{x} = -2$$

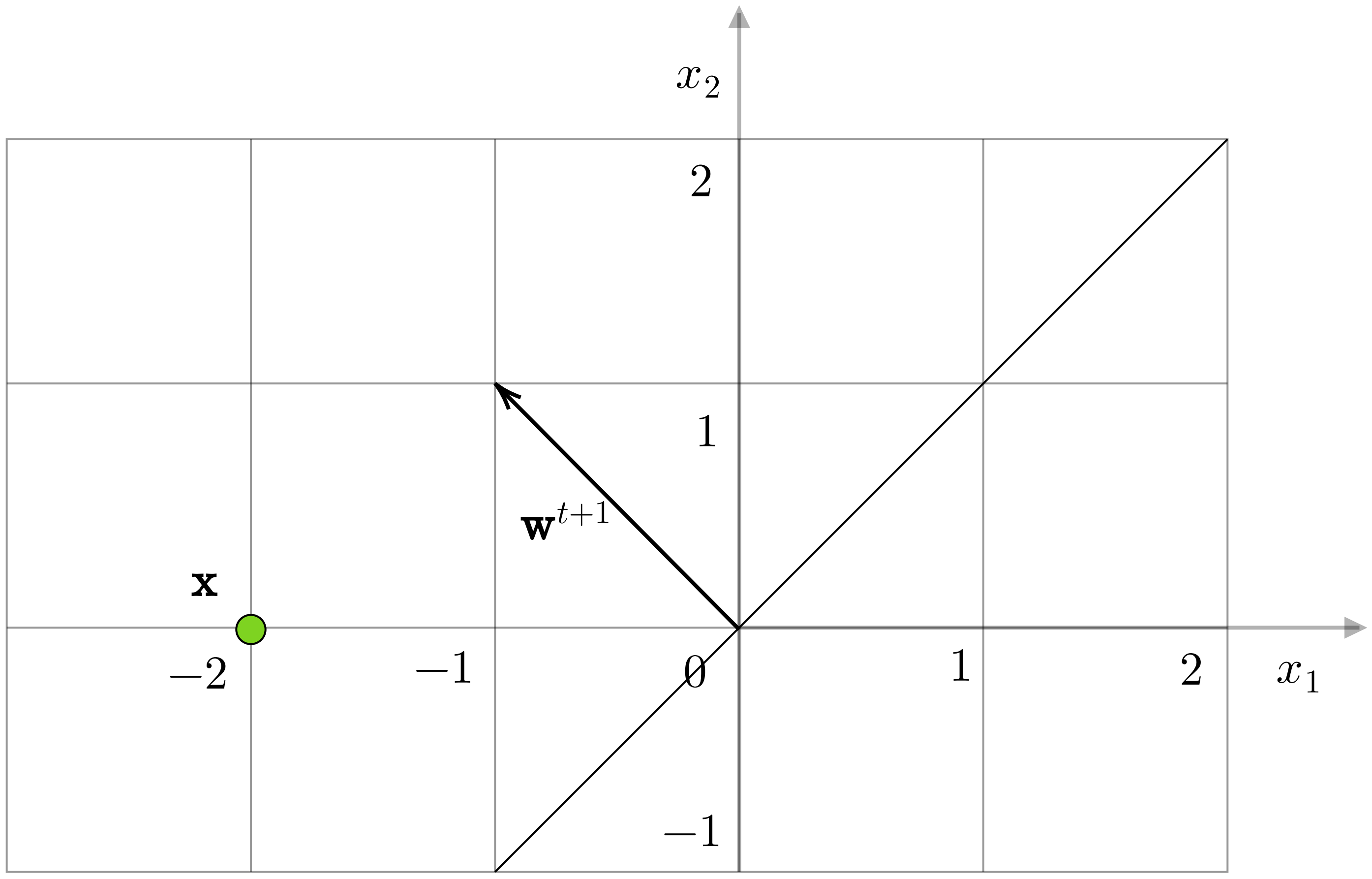
Actual = 1 Predicted = -1

$$\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}y$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \times \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Perceptron Update Rule



False Negative

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\mathbf{w}^{t\,T}\mathbf{x} = -2$$

Actual = 1 Predicted = -1

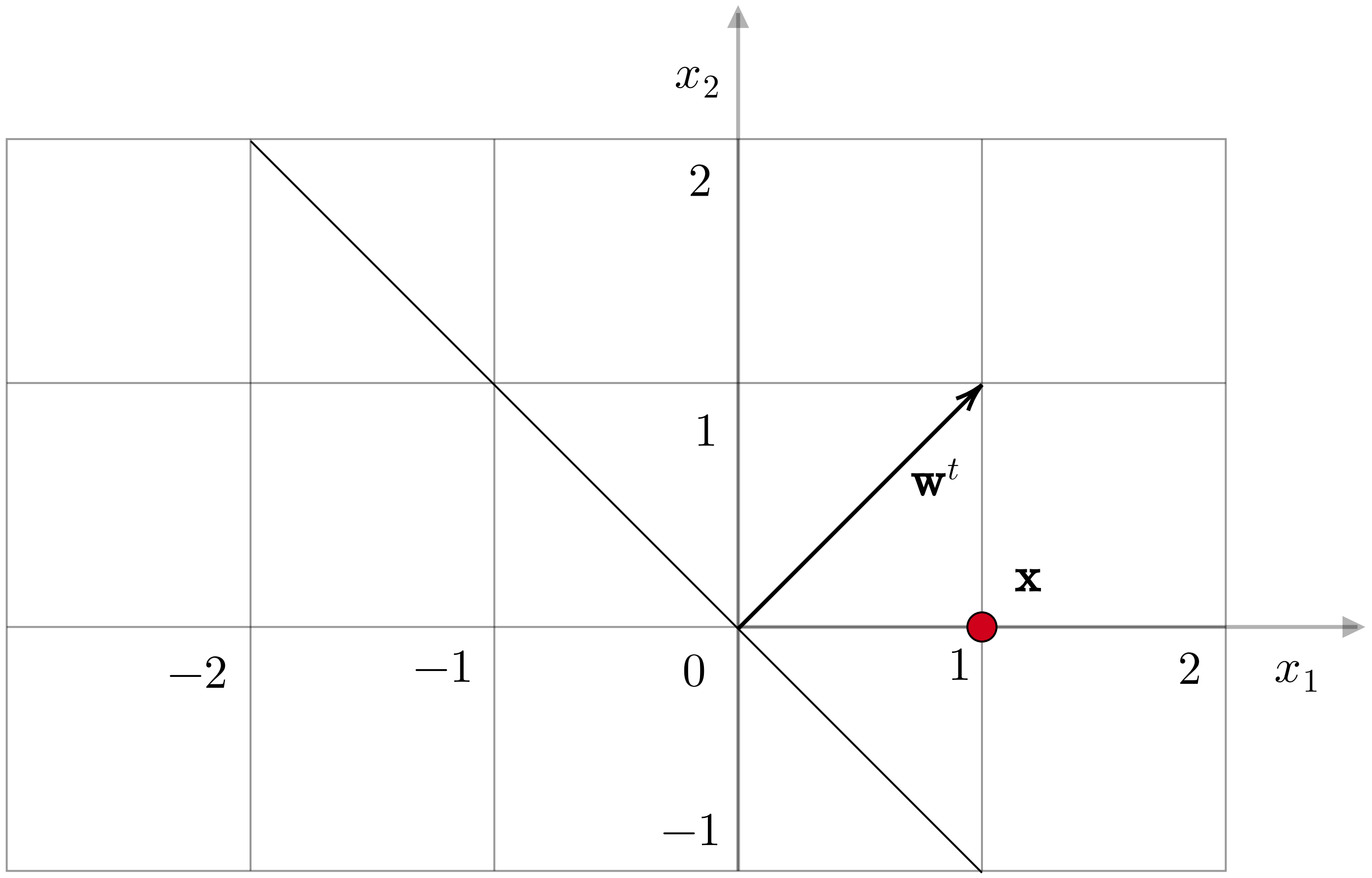
$$\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}y$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \times \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

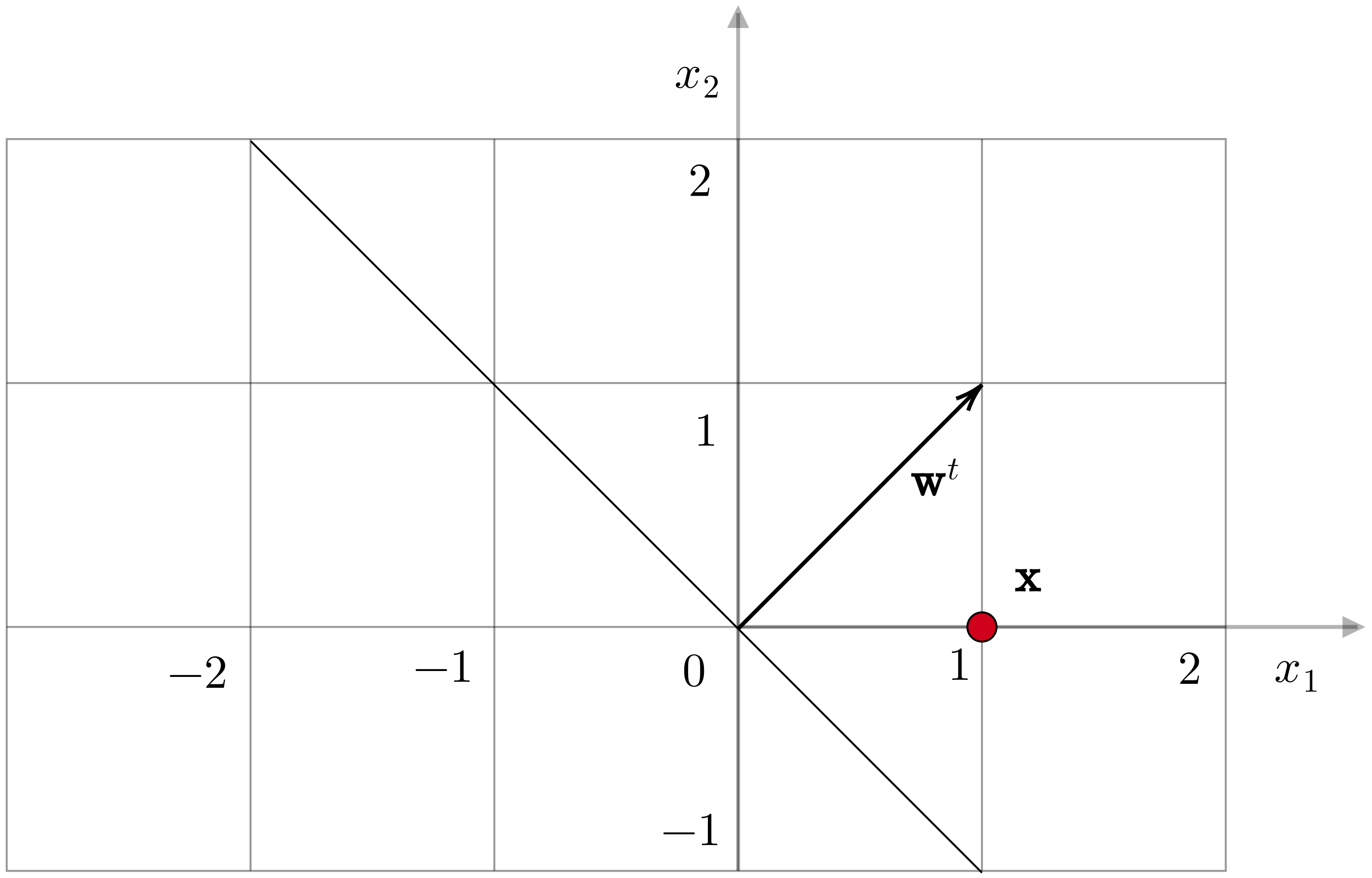
$$\mathbf{w}^{t+1\,T}\mathbf{x} = 2$$

Perceptron Update Rule



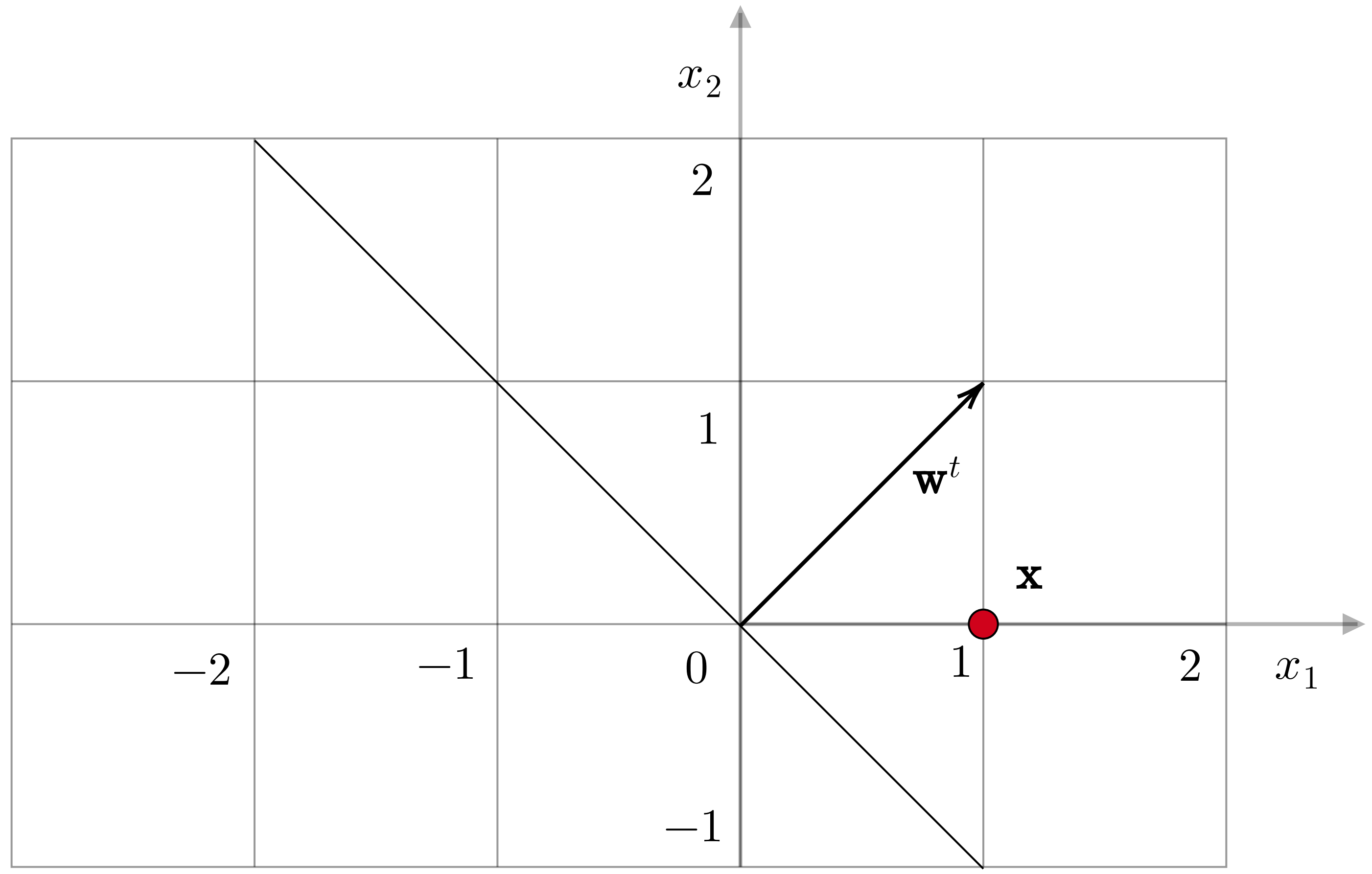
Perceptron Update Rule

False Positive



Perceptron Update Rule

False Positive

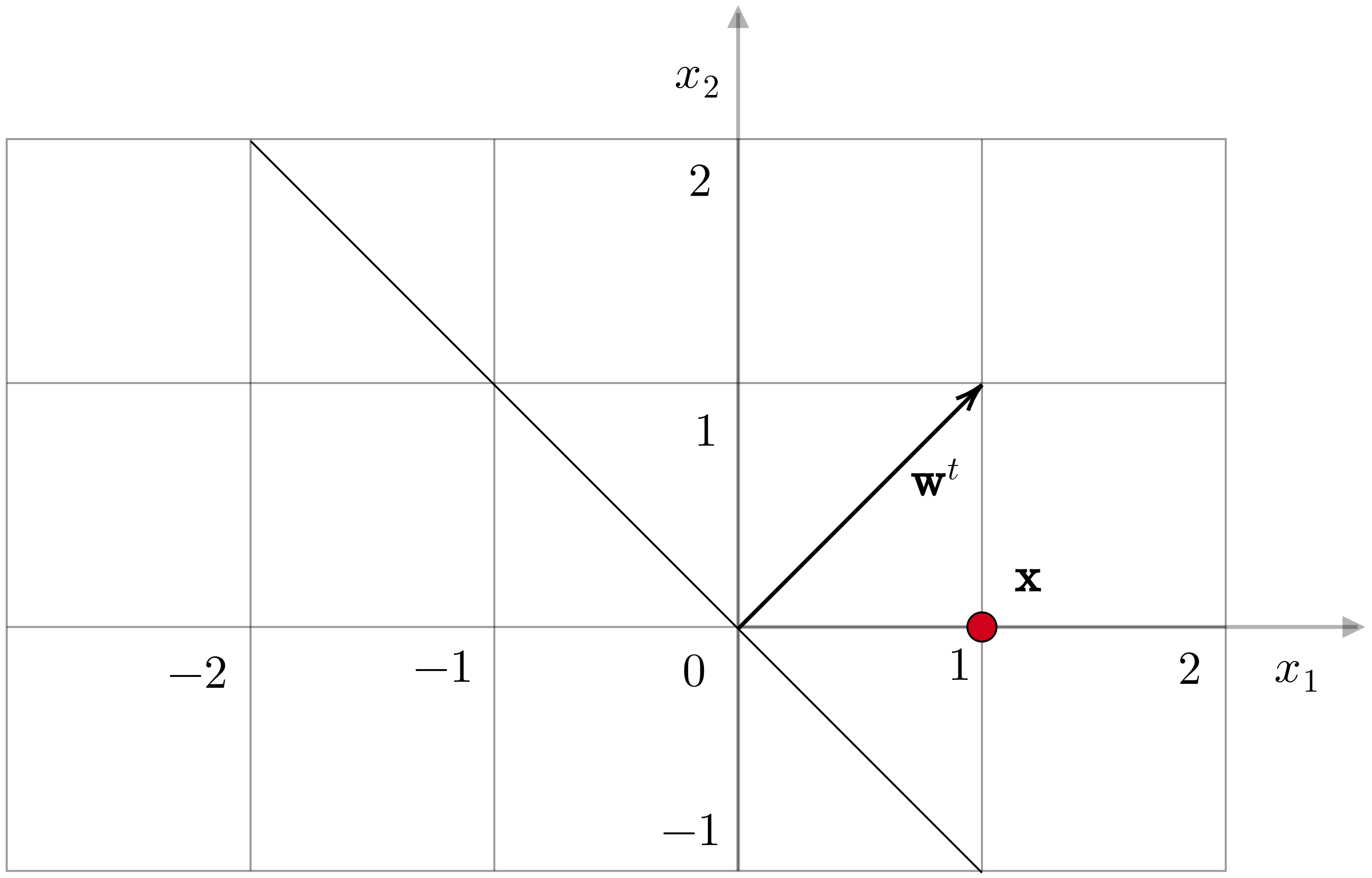


$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$${\mathbf{w}^t}^T \mathbf{x} = 1$$

Actual = -1 Predicted = 1

Perceptron Update Rule



False Positive

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{w}^{t^T}\mathbf{x} = 1$$

Actual = -1 Predicted = 1

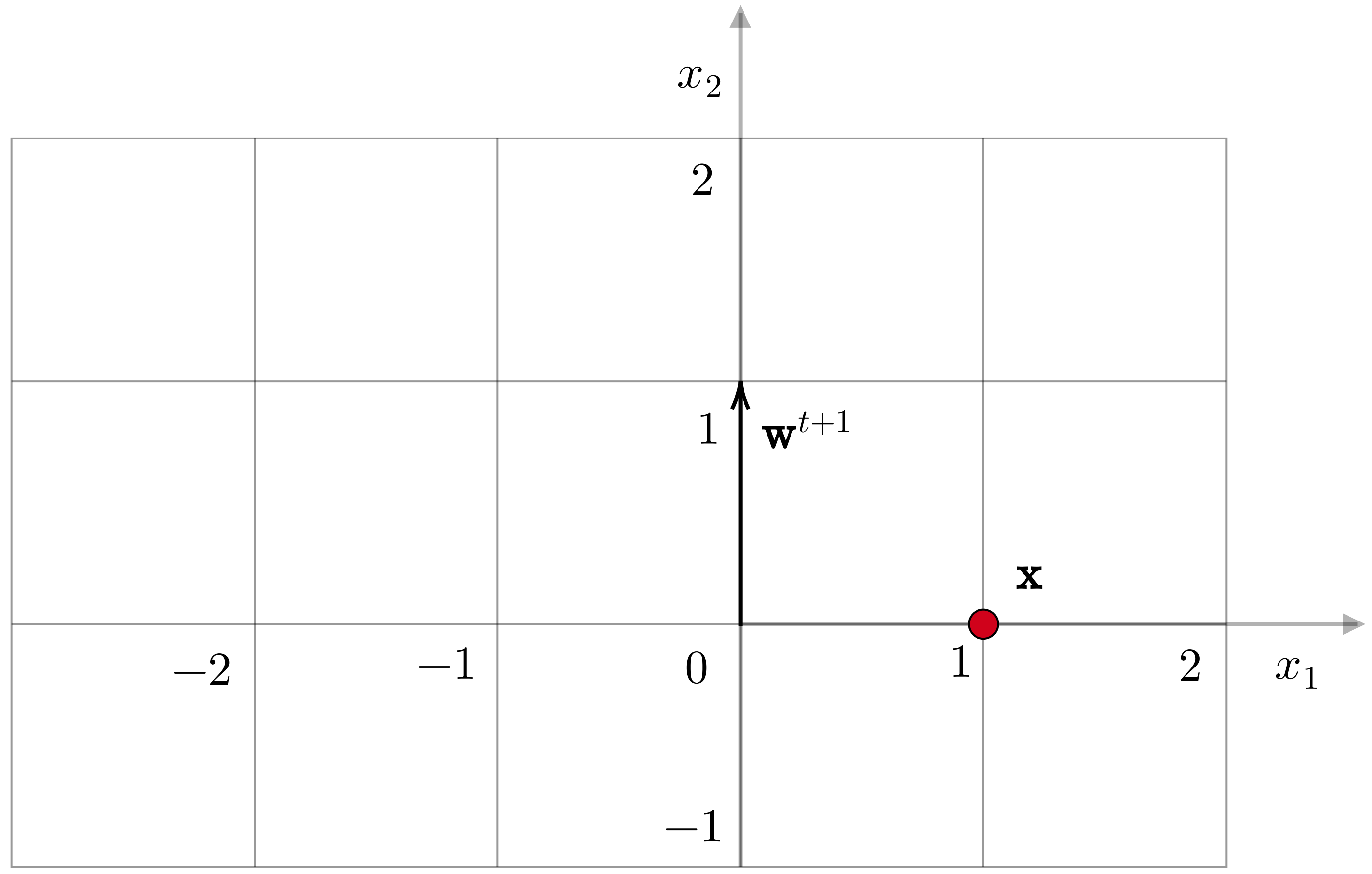
$$\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}y$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathbf{w}^{t+1^T}\mathbf{x} = 0$$

Perceptron Update Rule



False Positive

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$${\mathbf{w}^t}^T \mathbf{x} = 1$$

Actual = -1 Predicted = 1

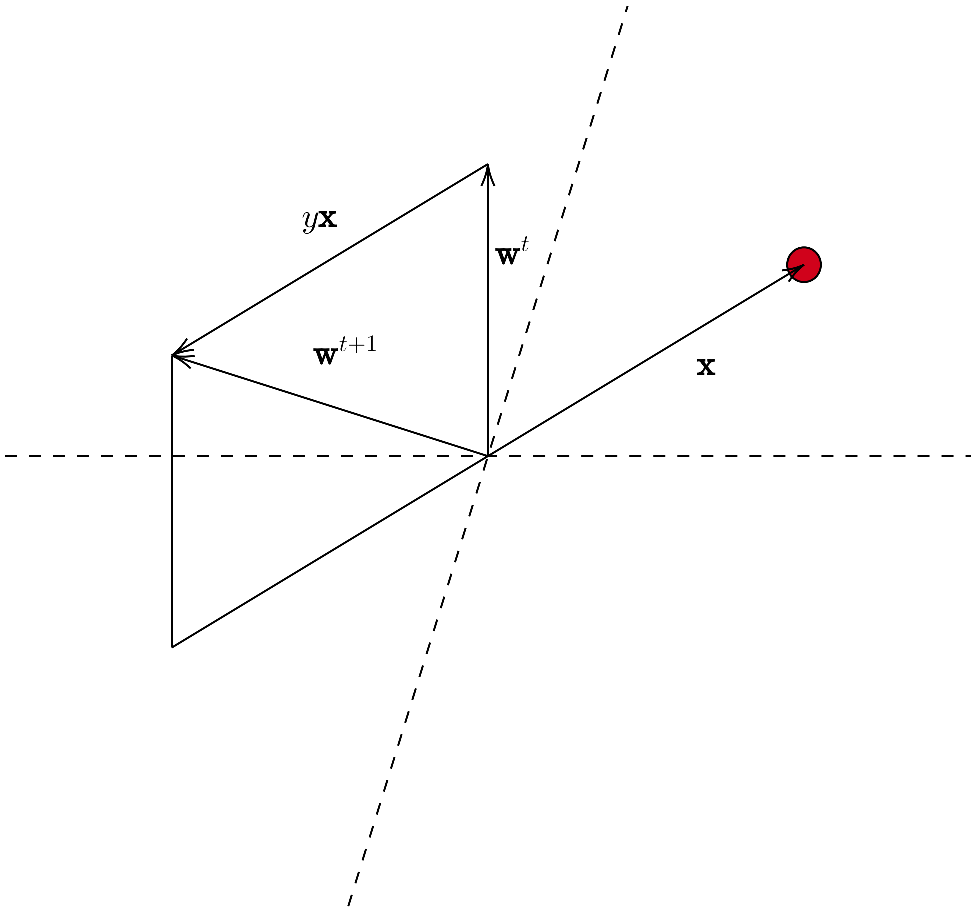
$$\mathbf{w}^{t+1} = \mathbf{w}^t + \mathbf{x}y$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

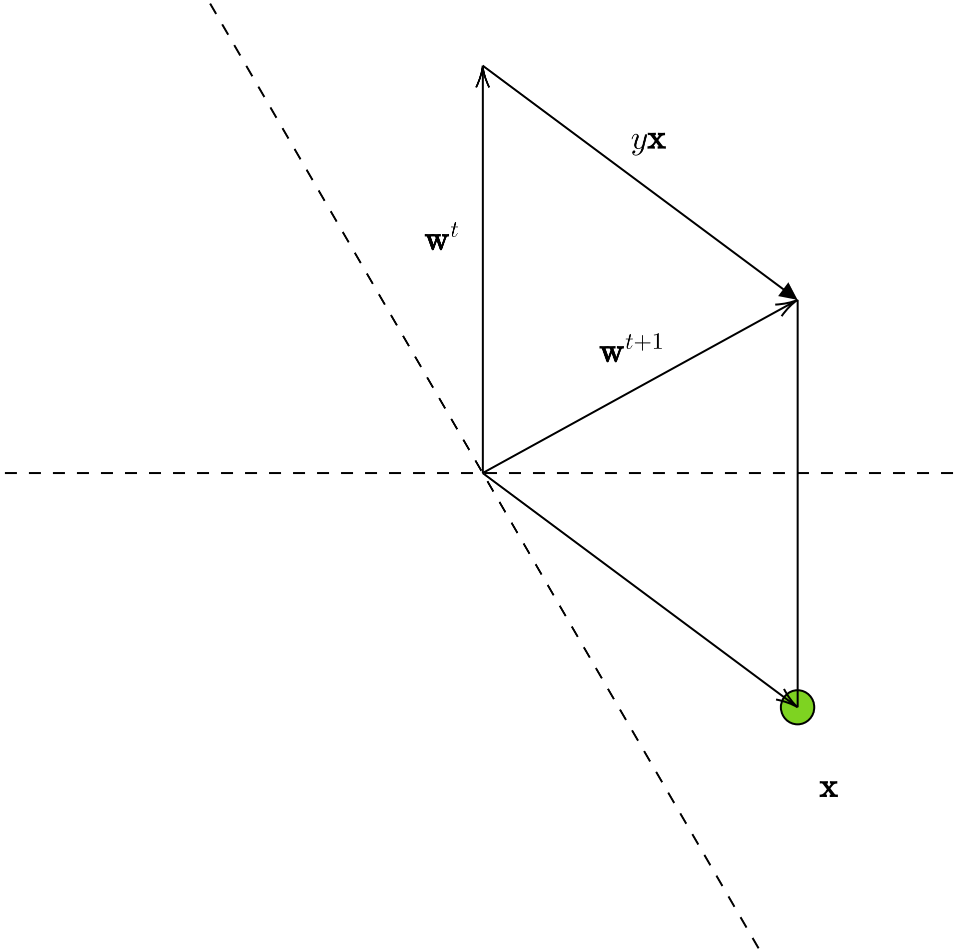
$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$${\mathbf{w}^{t+1}}^T \mathbf{x} = 0$$

Perceptron Update Rule



False Positive



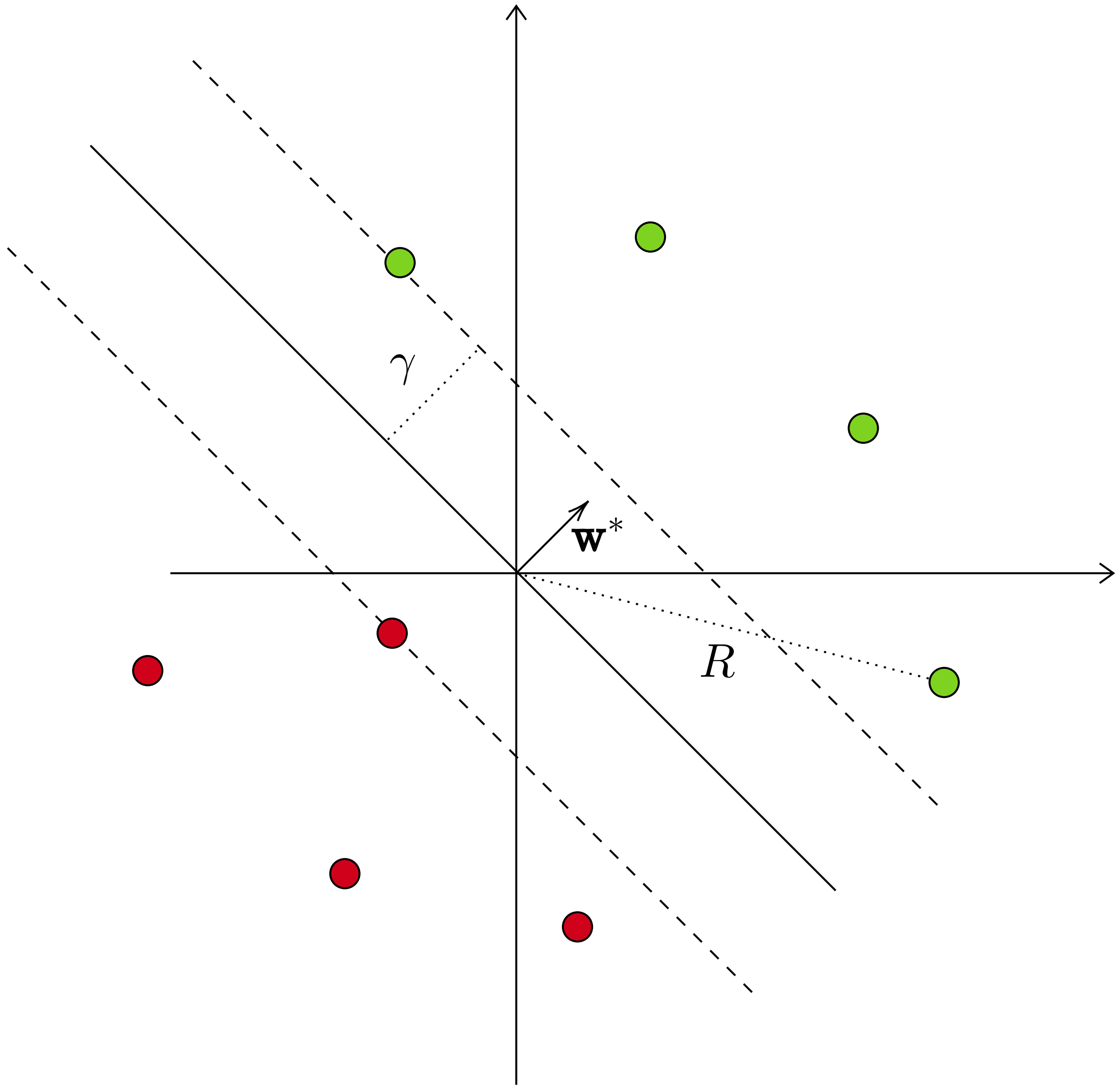
False Negative

Proof of Convergence

Assumptions

Proof of Convergence

Assumptions

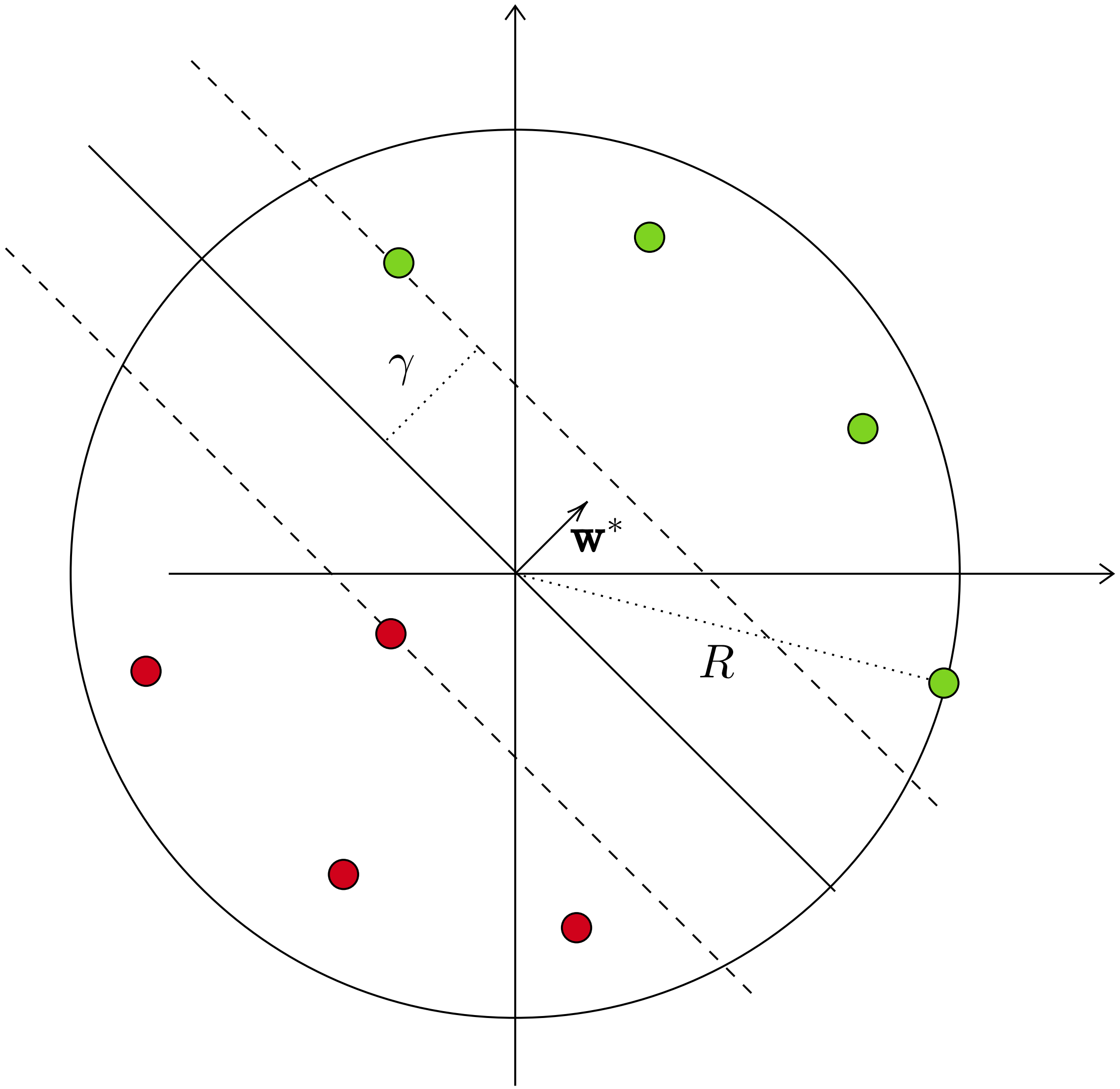


Linear Separability with γ-Margin

$$\left[\left(\mathbf{w}^*\right)^T \mathbf{x}_i\right] y_i \geqslant \gamma \qquad ||\mathbf{w}^*|| = 1$$

with $\gamma > 0$

Proof of Convergence



Assumptions

Linear Separability with γ-Margin

$$\left[\left(\mathbf{w}^*\right)^T \mathbf{x}_i\right] y_i \geqslant \gamma \qquad ||\mathbf{w}^*|| = 1$$

with $\gamma > 0$

Radius Bound

$$||\mathbf{x}||^2 \leqslant R^2$$

Proof of Convergence

Upper Bound

$$||\mathbf{w}^t||^2 = \left(\mathbf{w}^t\right)^T \mathbf{w}^t$$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)
\end{aligned}
$$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x}
\end{aligned}
$$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

$$||\mathbf{w}^t||^2 \leqslant ||\mathbf{w}^{t-1}||^2 + R^2$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

$$
\begin{aligned}
||\mathbf{w}^t||^2 &\leqslant ||\mathbf{w}^{t-1}||^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-2}||^2 + R^2 + R^2
\end{aligned}
$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

$$
\begin{aligned}
||\mathbf{w}^t||^2 &\leqslant ||\mathbf{w}^{t-1}||^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-2}||^2 + R^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-3}||^2 + R^2 + R^2 + R^2
\end{aligned}
$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$
\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}
$$

$$
\begin{aligned}
||\mathbf{w}^t||^2 &\leqslant ||\mathbf{w}^{t-1}||^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-2}||^2 + R^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-3}||^2 + R^2 + R^2 + R^2 \\
&\leqslant \qquad \vdots \\
&\leqslant ||\mathbf{w}^{t-t}||^2 + tR^2 \\
&= ||\mathbf{w}^{0}||^2 + tR^2
\end{aligned}
$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}$$

$$\begin{aligned}
||\mathbf{w}^t||^2 &\leqslant ||\mathbf{w}^{t-1}||^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-2}||^2 + R^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-3}||^2 + R^2 + R^2 + R^2 \\
&\leqslant \qquad \vdots \\
&\leqslant ||\mathbf{w}^{t-t}||^2 + tR^2 \\
&= ||\mathbf{w}^0||^2 + tR^2 \\
&= t \cdot R^2
\end{aligned}$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

Proof of Convergence

Upper Bound

$$\begin{aligned}
||\mathbf{w}^t||^2 &= \left(\mathbf{w}^t\right)^T \mathbf{w}^t \\
&= \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= ||\mathbf{w}^{t-1}||^2 + y^2 ||\mathbf{x}||^2 + 2y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \\
&\leqslant ||\mathbf{w}^{t-1}||^2 + R^2
\end{aligned}$$

$$\begin{aligned}
||\mathbf{w}^t||^2 &\leqslant ||\mathbf{w}^{t-1}||^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-2}||^2 + R^2 + R^2 \\
&\leqslant ||\mathbf{w}^{t-3}||^2 + R^2 + R^2 + R^2 \\
&\leqslant \qquad \vdots \\
&\leqslant ||\mathbf{w}^{t-t}||^2 + tR^2 \\
&= ||\mathbf{w}^0||^2 + tR^2 \\
&= t \cdot R^2
\end{aligned}$$

Facts

- $y^2 = 1$
- $||\mathbf{x}||^2 \leqslant R^2$
- $y \cdot \left(\mathbf{w}^{t-1}\right)^T \mathbf{x} \leqslant 0$
- $||\mathbf{w}^0|| = 0$

$$\boxed{||\mathbf{w}^t||^2 \leqslant t \cdot R^2}$$

Proof of Convergence

Lower Bound

$$\left(\mathbf{w}^*\right)^T \mathbf{w}^t = \left(\mathbf{w}^*\right)^T (\mathbf{w}^{t-1} + y \cdot \mathbf{x})$$

Proof of Convergence

Lower Bound

$$\left(\mathbf{w}^*\right)^T \mathbf{w}^t = \left(\mathbf{w}^*\right)^T (\mathbf{w}^{t-1} + y \cdot \mathbf{x})$$

$$= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x}$$

Proof of Convergence

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$\left(\mathbf{w}^*\right)^T \mathbf{w}^t = \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)$$

$$= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x}$$

$$\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma$$

$$\left(\mathbf{w}^*\right)^T \mathbf{w}^t \geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-2} + \gamma + \gamma
\end{aligned}
$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T (\mathbf{w}^{t-1} + y \cdot \mathbf{x}) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-2} + \gamma + \gamma \\
&\vdots \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{0} + t\gamma
\end{aligned}
$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T (\mathbf{w}^{t-1} + y \cdot \mathbf{x}) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-2} + \gamma + \gamma \\
&\vdots \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{0} + t\gamma \\
&= t\gamma
\end{aligned}
$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &= \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right) \\
&= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathbf{w}^*\right)^T \mathbf{w}^t &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-2} + \gamma + \gamma \\
&\vdots \\
&\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{0} + t\gamma \\
&= t\gamma
\end{aligned}
$$

Cauchy-Schwartz

$$
\begin{aligned}
||\mathbf{w}^*|| \cdot ||\mathbf{w}^t|| &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^t \\
||\mathbf{w}^t|| &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^t
\end{aligned}
$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$\left(\mathbf{w}^*\right)^T \mathbf{w}^t = \left(\mathbf{w}^*\right)^T \left(\mathbf{w}^{t-1} + y \cdot \mathbf{x}\right)$$

$$= \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x}$$

$$\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma$$

$$\begin{aligned}\left(\mathbf{w}^*\right)^T \mathbf{w}^t &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-1} + \gamma \\ &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{t-2} + \gamma + \gamma \\ &\vdots \\ &\geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^{0} + t\gamma \\ &= t\gamma\end{aligned}$$

Cauchy-Schwartz

$$||\mathbf{w}^*|| \cdot ||\mathbf{w}^t|| \geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^t$$

$$||\mathbf{w}^t|| \geqslant \left(\mathbf{w}^*\right)^T \mathbf{w}^t$$

Facts

- $y \cdot \left(\mathbf{w}^*\right)^T \mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^0|| = 0$
- $||\mathbf{w}^*|| = 1$

Proof of Convergence

Lower Bound

$$
\begin{aligned}
\left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t} &= \left(\mathbf{w}^{*}\right)^{T}\left(\mathbf{w}^{t-1} + y\cdot\mathbf{x}\right) \\
&= \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t-1} + y\cdot\left(\mathbf{w}^{*}\right)^{T}\mathbf{x} \\
&\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t-1} + \gamma
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t} &\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t-1} + \gamma \\
&\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t-2} + \gamma + \gamma \\
&\vdots \\
&\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{0} + t\gamma \\
&= t\gamma
\end{aligned}
$$

Cauchy-Schwartz

$$
\begin{aligned}
||\mathbf{w}^{*}||\cdot||\mathbf{w}^{t}|| &\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t} \\
||\mathbf{w}^{t}|| &\geqslant \left(\mathbf{w}^{*}\right)^{T}\mathbf{w}^{t}
\end{aligned}
$$

Facts

- $y\cdot\left(\mathbf{w}^{*}\right)^{T}\mathbf{x} \geqslant \gamma$
- $||\mathbf{w}^{0}|| = 0$
- $||\mathbf{w}^{*}|| = 1$

$$\boxed{||\mathbf{w}^{t}||^{2} \geqslant t^{2}\cdot\gamma^{2}}$$

Proof of Convergence

$$t^2 \cdot \gamma^2 \leqslant ||\mathbf{w}^t||^2 \leqslant t \cdot R^2$$

$$t^2\gamma^2 \leqslant t \cdot R^2$$

$$t \leqslant \frac{R^2}{\gamma^2}$$

$$\boxed{t \leqslant \frac{R^2}{\gamma^2}}$$

Radius-Margin Bound

Proof of Convergence

$$t^2 \cdot \gamma^2 \leqslant ||\mathbf{w}^t||^2 \leqslant t \cdot R^2$$

$$t^2\gamma^2 \leqslant t \cdot R^2$$

$$t \leqslant \frac{R^2}{\gamma^2}$$

$$\boxed{t \leqslant \frac{R^2}{\gamma^2}}$$

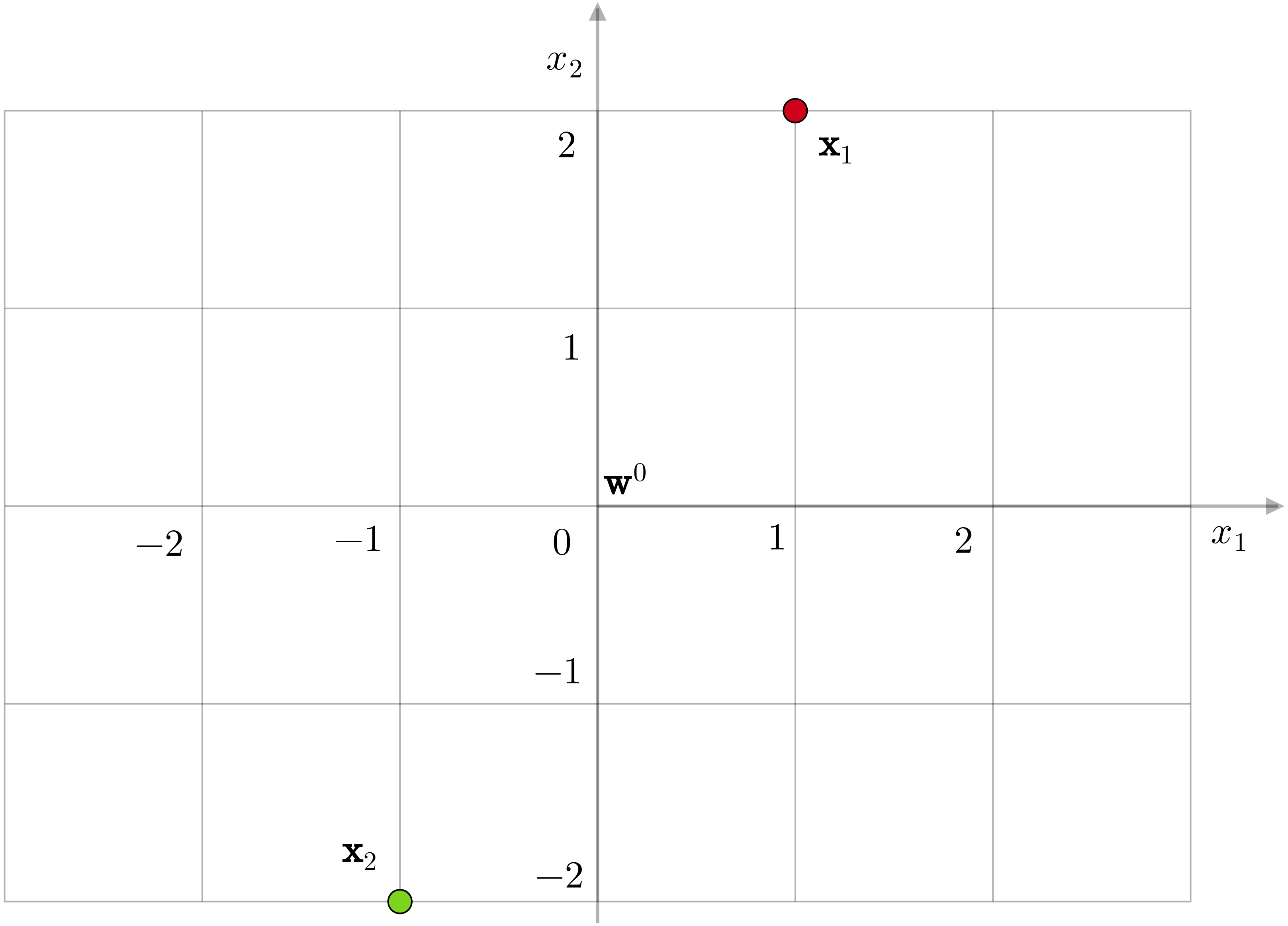
Radius-Margin Bound

Bound and Interpretation

- Number of updates to the weight vector is at most $\frac{R^2}{\gamma^2}$
- Number of mistakes seen by the perceptron is at most $\frac{R^2}{\gamma^2}$

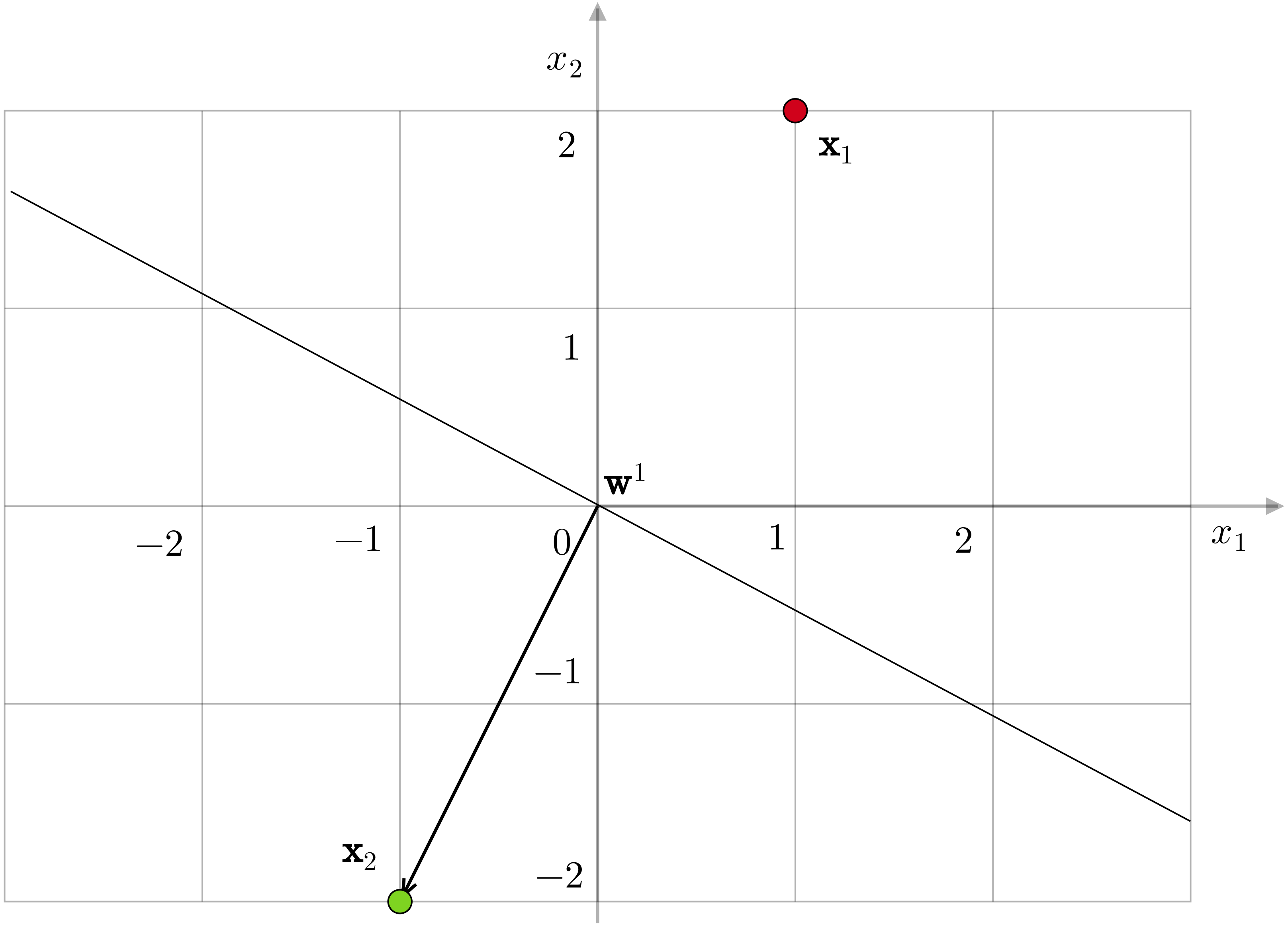
Example-1

$$\mathbf{X} = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$



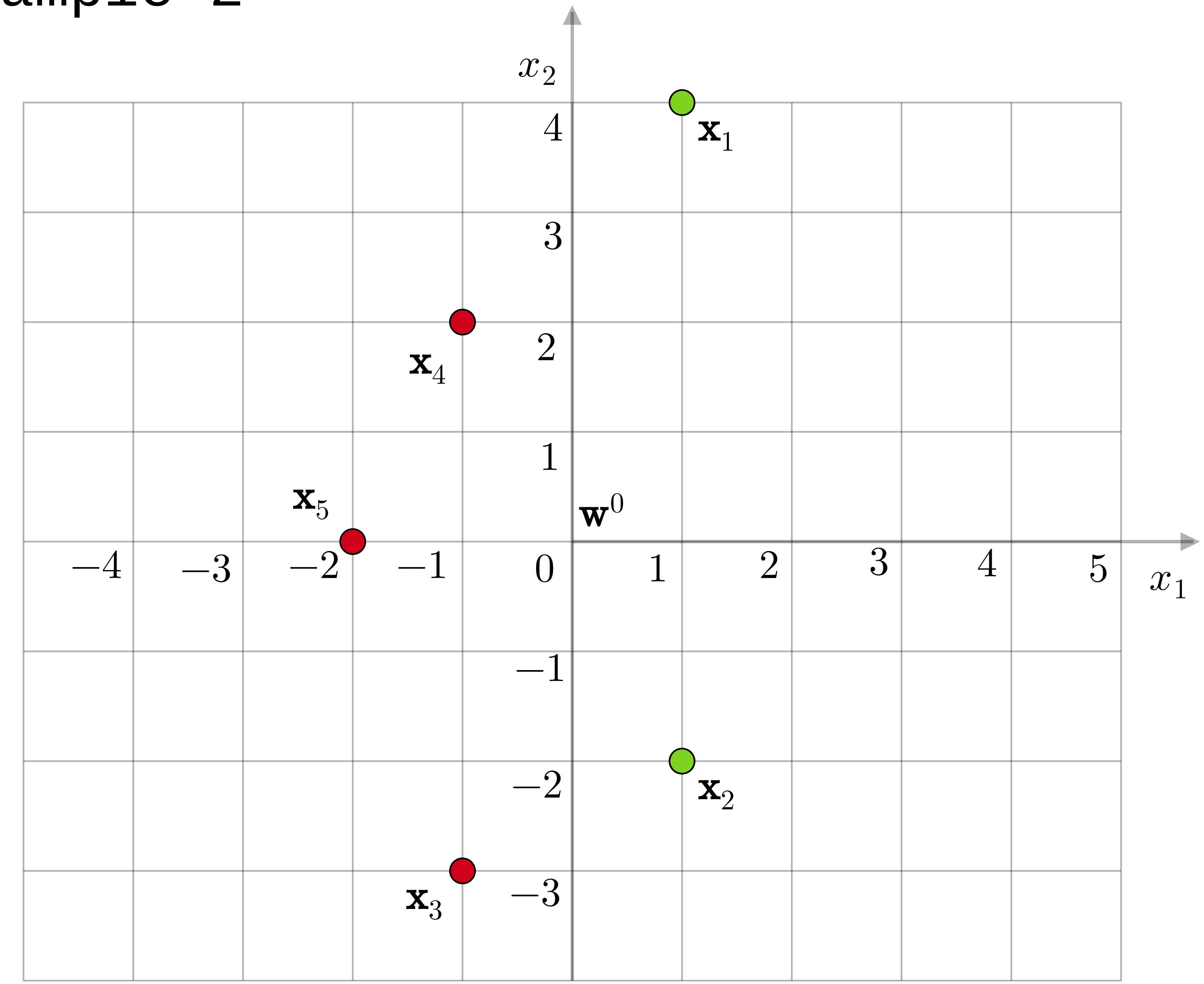
Example-1, Update-1

$$\mathbf{X} = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$



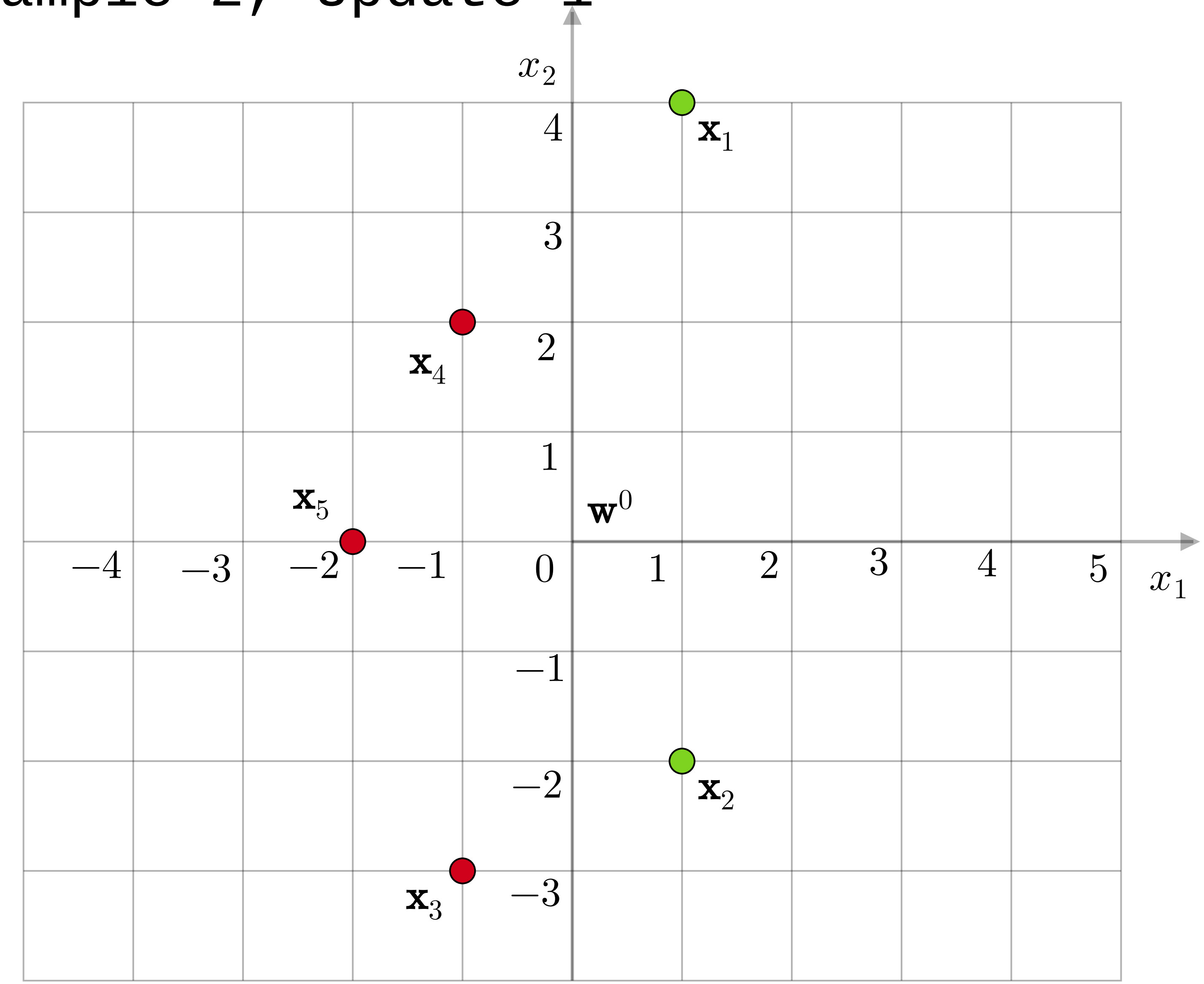
$$\begin{aligned} \mathbf{w}^1 &= \mathbf{w}^0 + \mathbf{x}_1 y_1 \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ -2 \end{bmatrix} \end{aligned}$$

Example-2



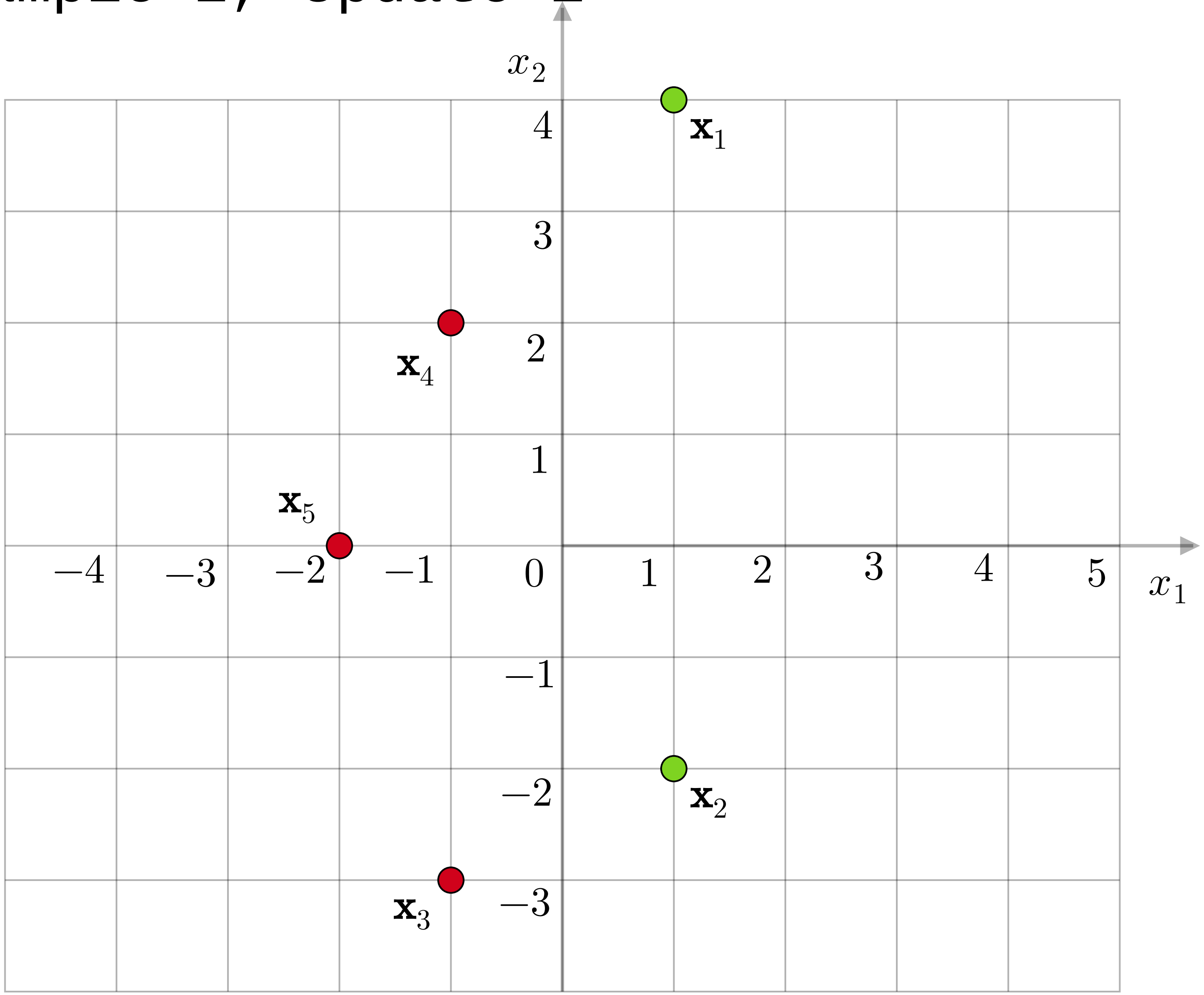
$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

Example-2, Update-1



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

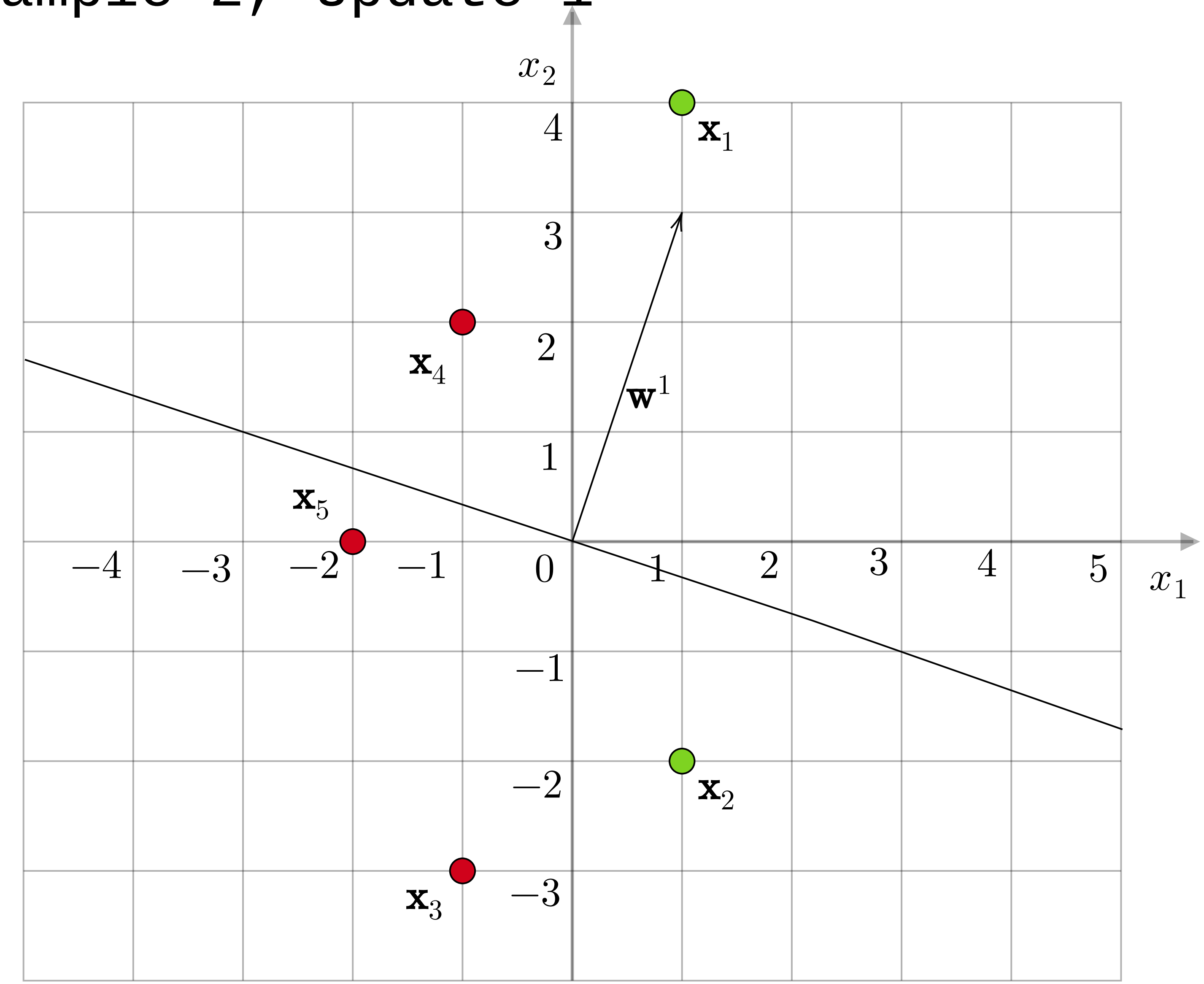
Example-2, Update-1



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$$\mathbf{w}^1 = \mathbf{w}^0 + \mathbf{x}_3 y_3$$

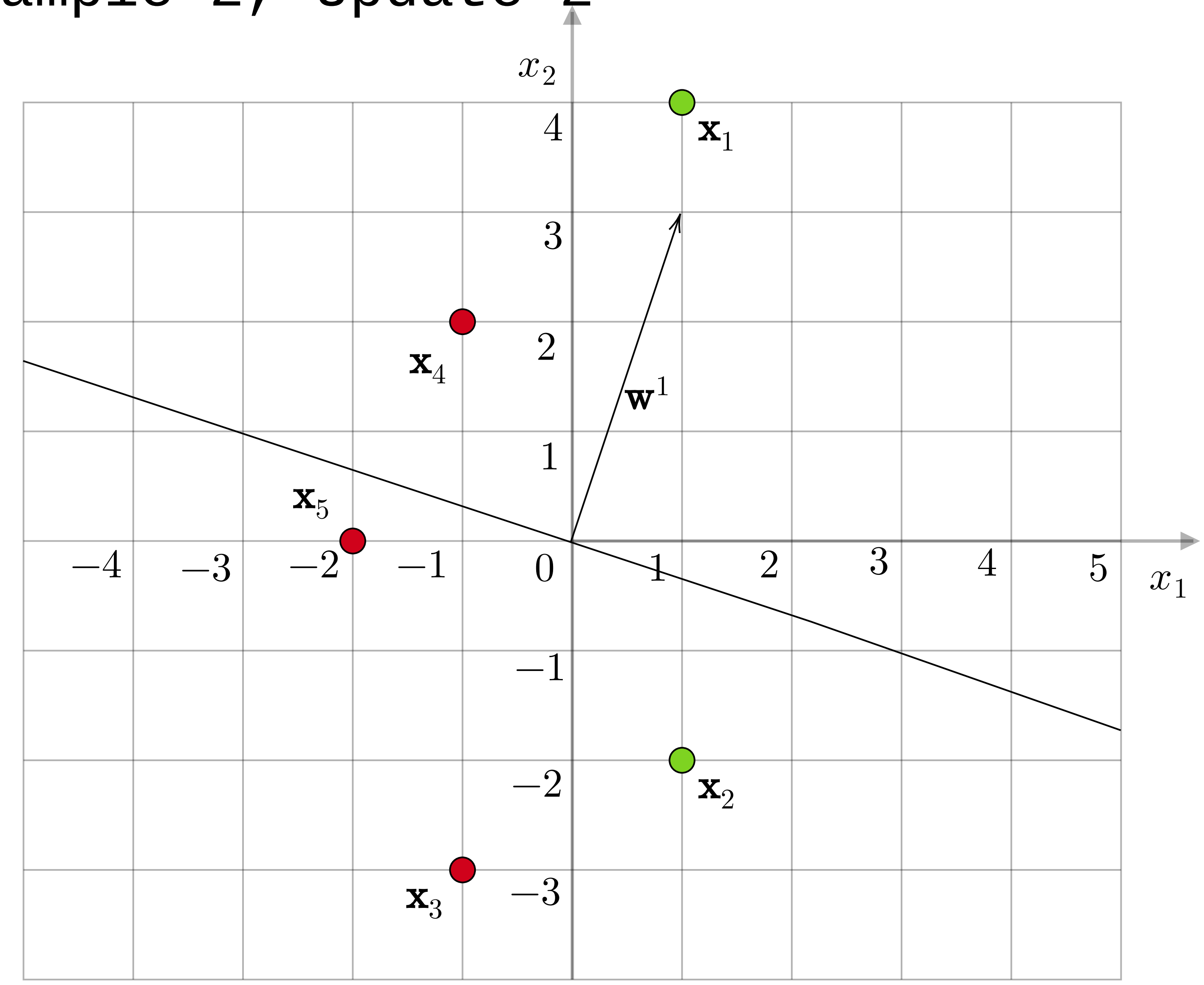
Example-2, Update-1



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

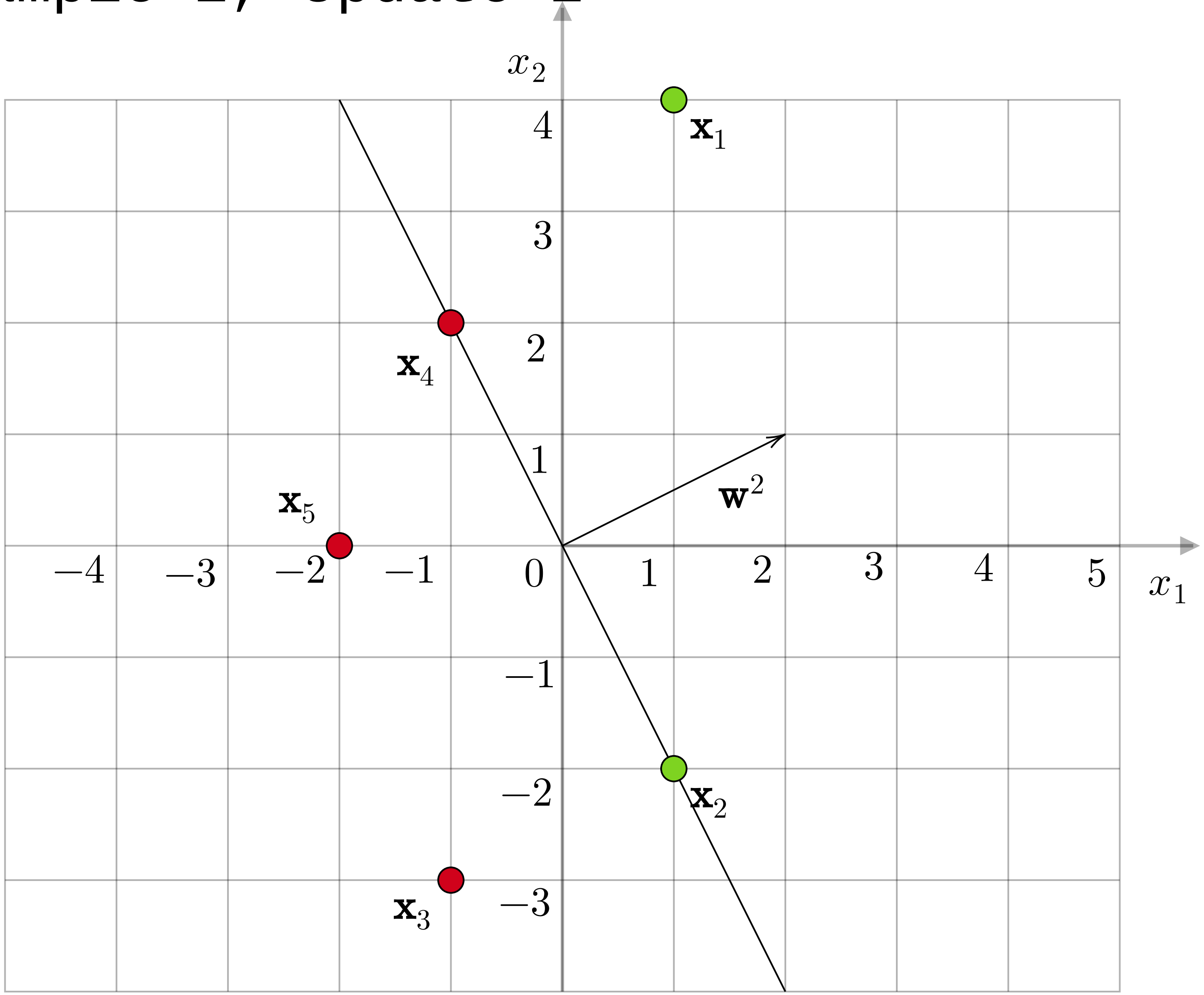
$$\begin{aligned} \mathbf{w}^1 &= \mathbf{w}^0 + \mathbf{x}_3 y_3 \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ -3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 3 \end{bmatrix} \end{aligned}$$

Example-2, Update-2



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

Example-2, Update-2



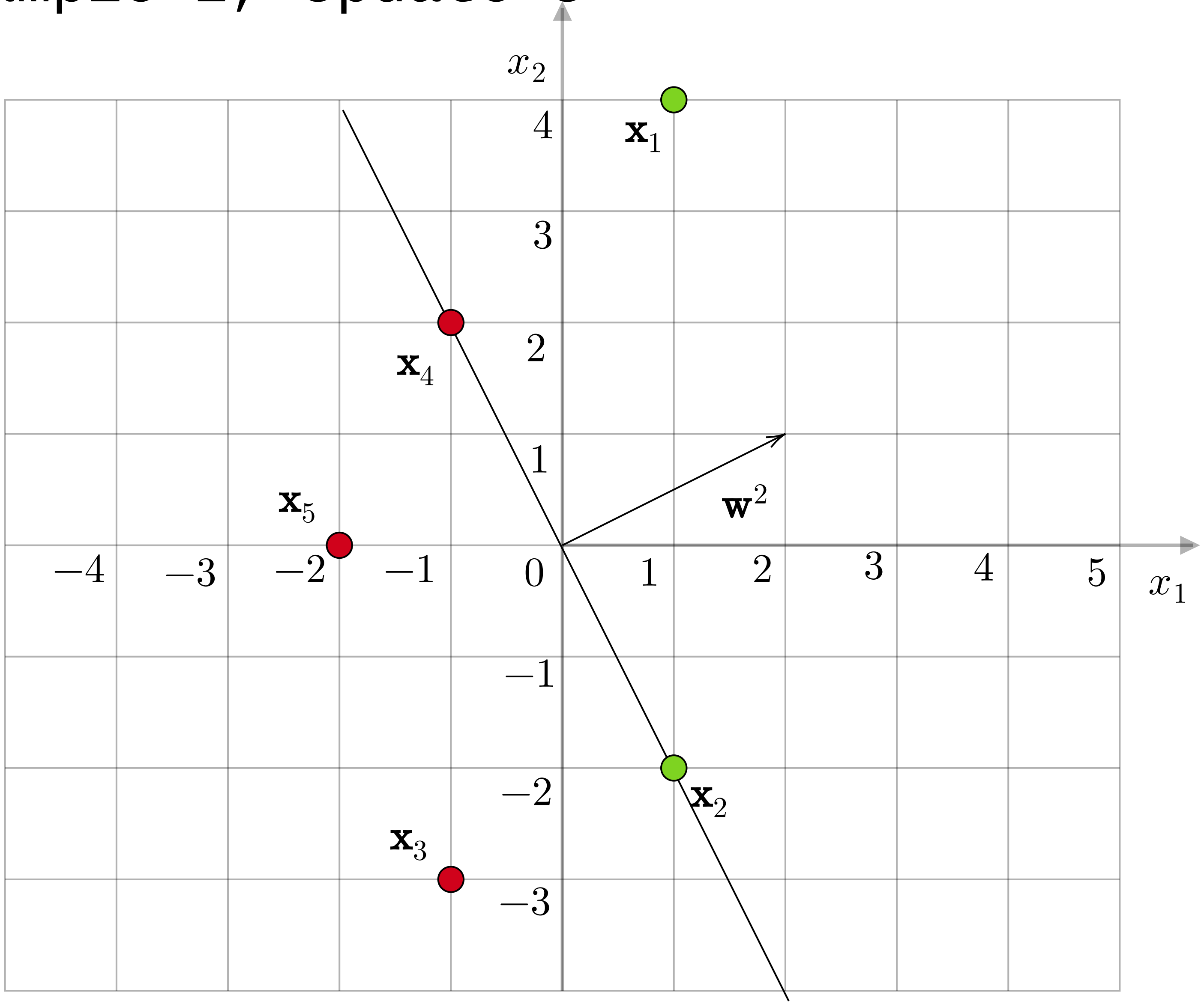
$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$$\mathbf{w}^2 = \mathbf{w}^1 + \mathbf{x}_4 y_4$$

$$= \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

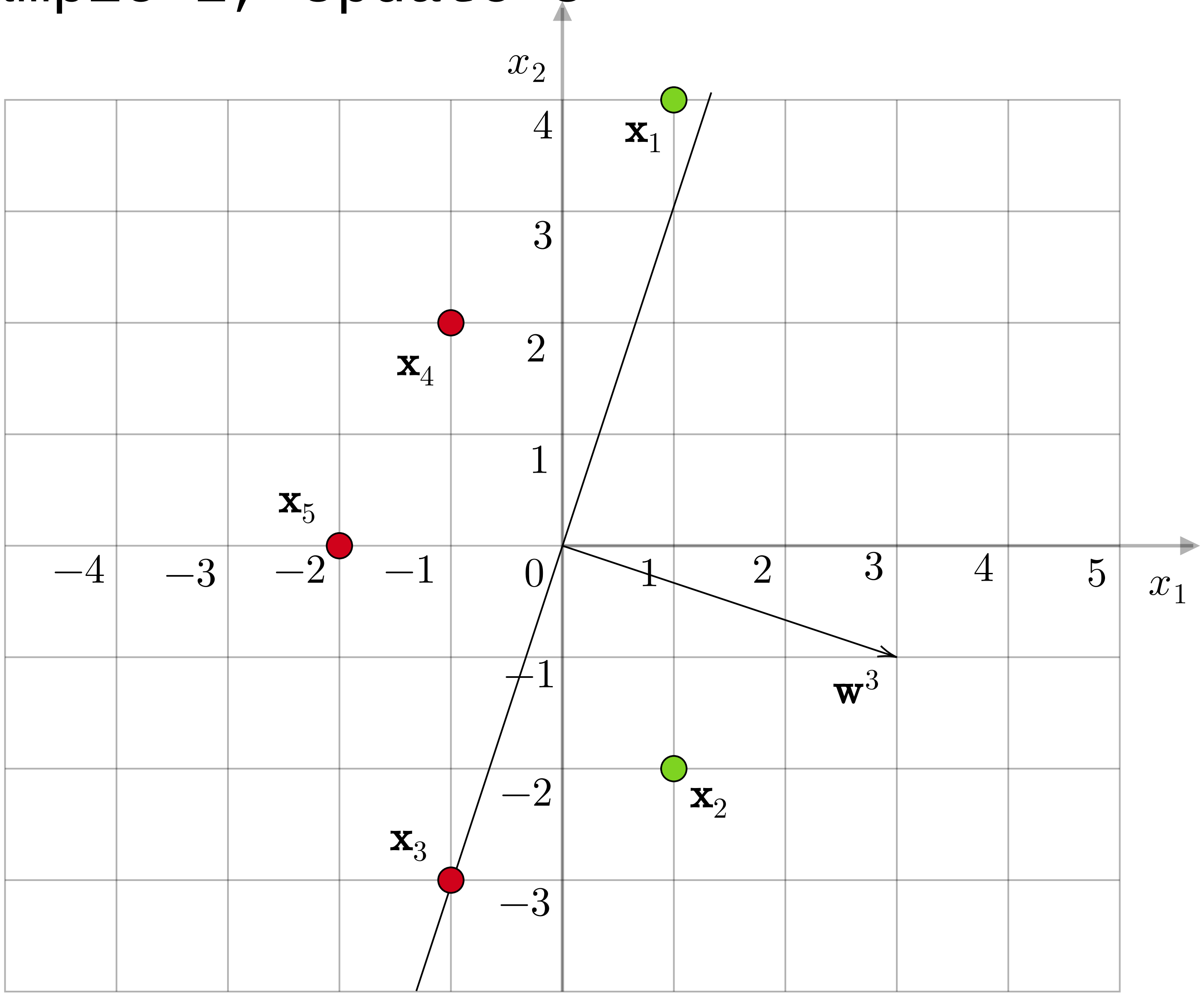
$$= \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Example-2, Update-3



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

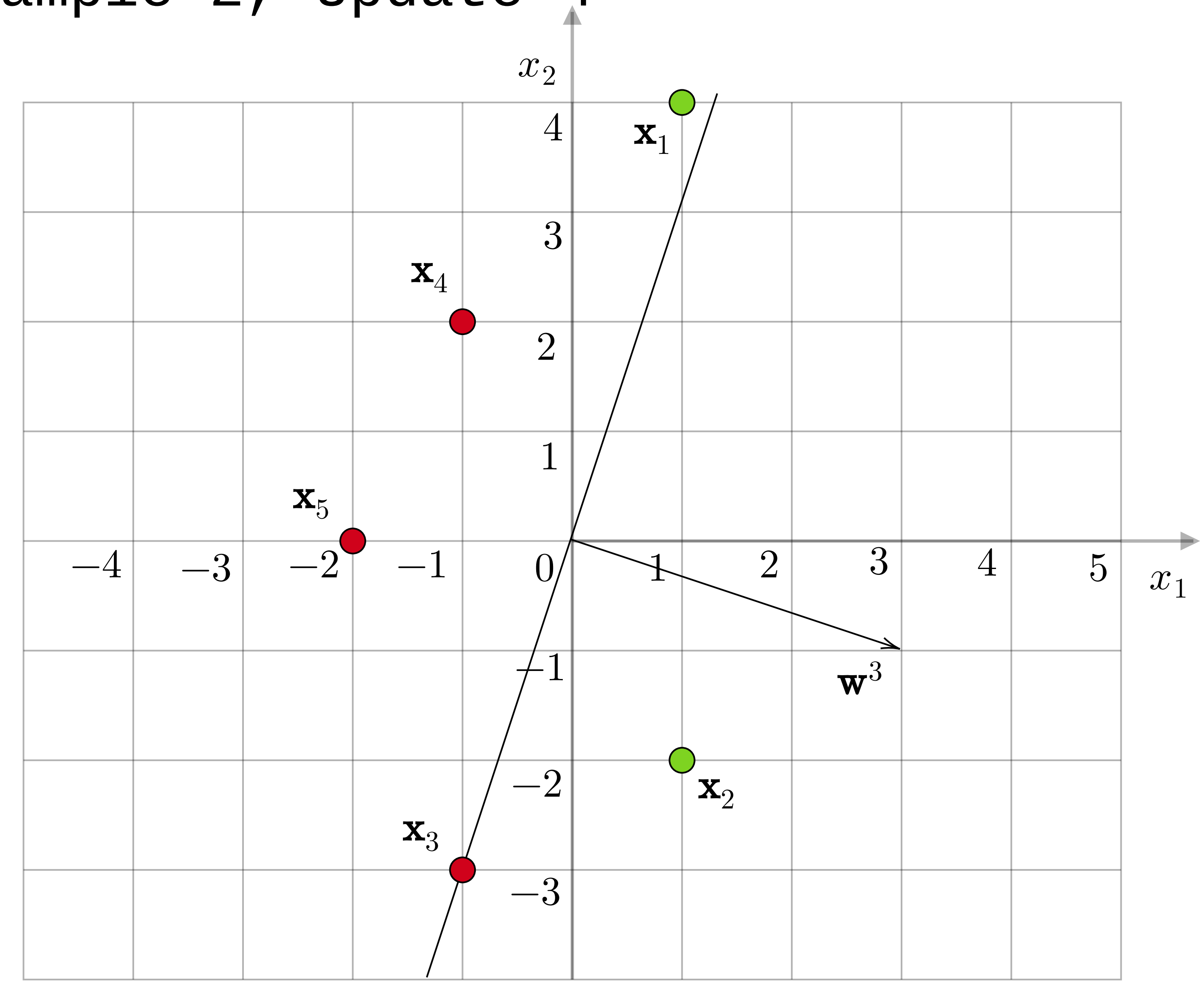
Example-2, Update-3



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

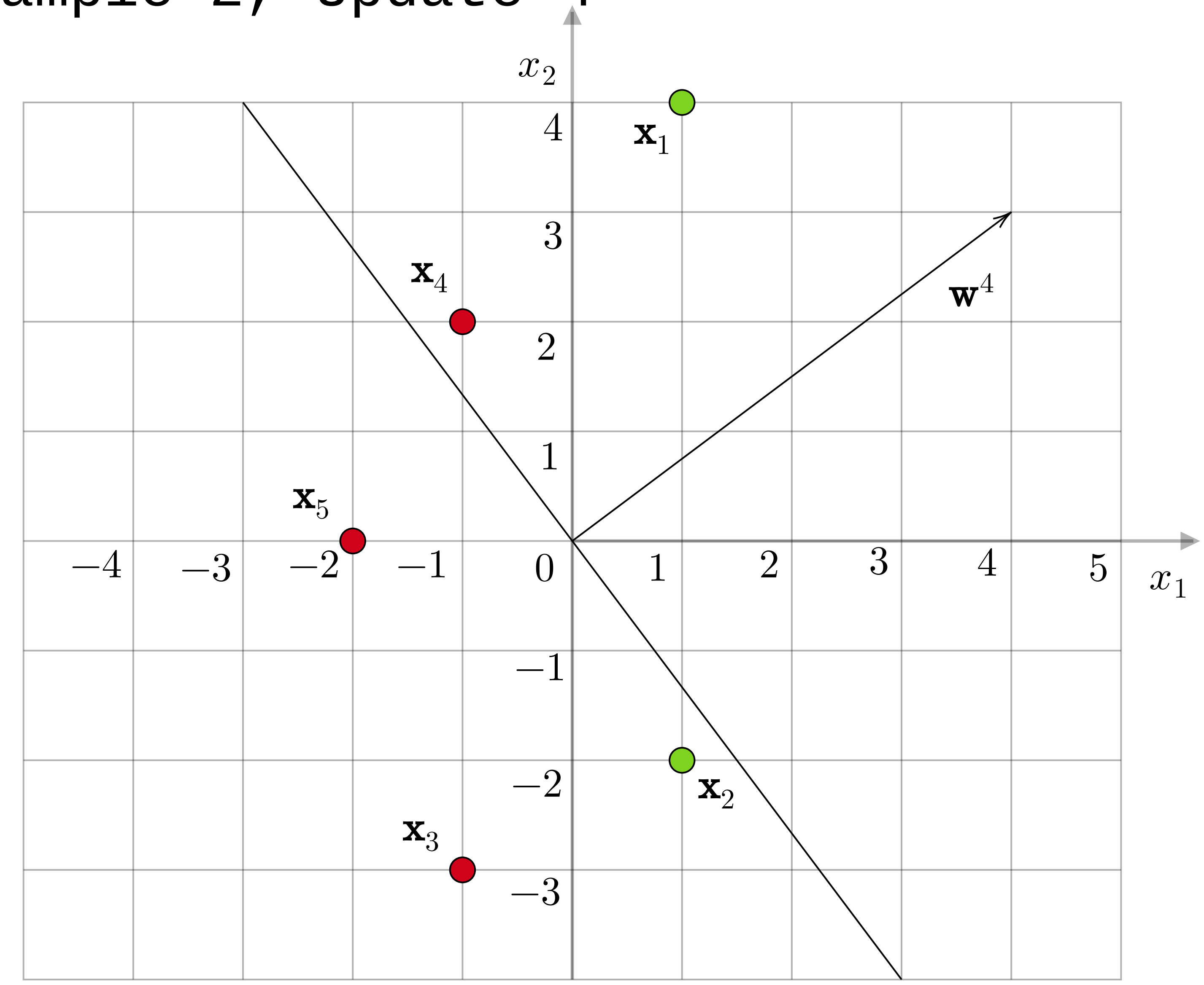
$$\begin{aligned} \mathbf{w}^3 &= \mathbf{w}^2 + \mathbf{x}_4 y_4 \\ &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ -1 \end{bmatrix} \end{aligned}$$

Example-2, Update-4



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

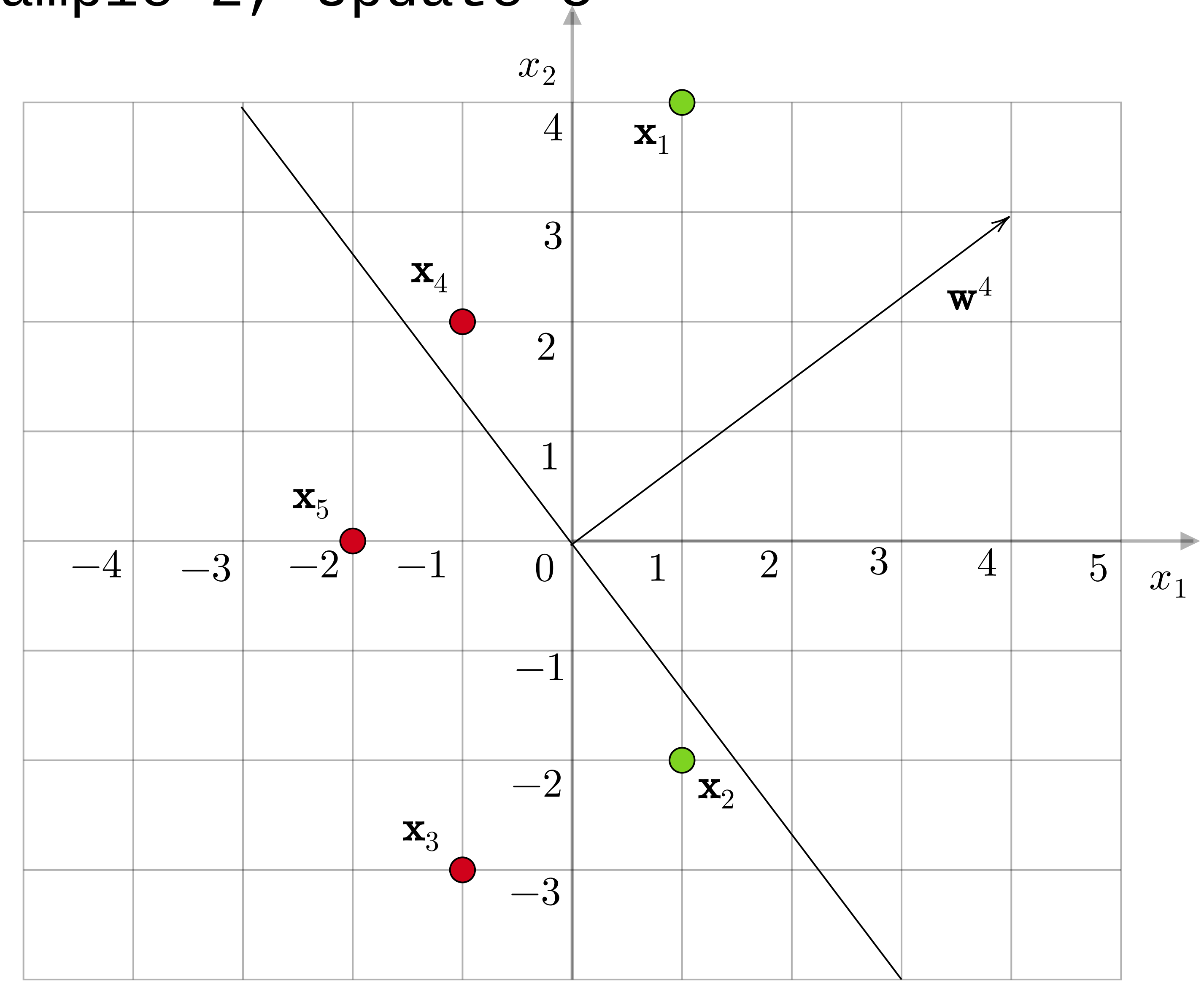
Example-2, Update-4



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

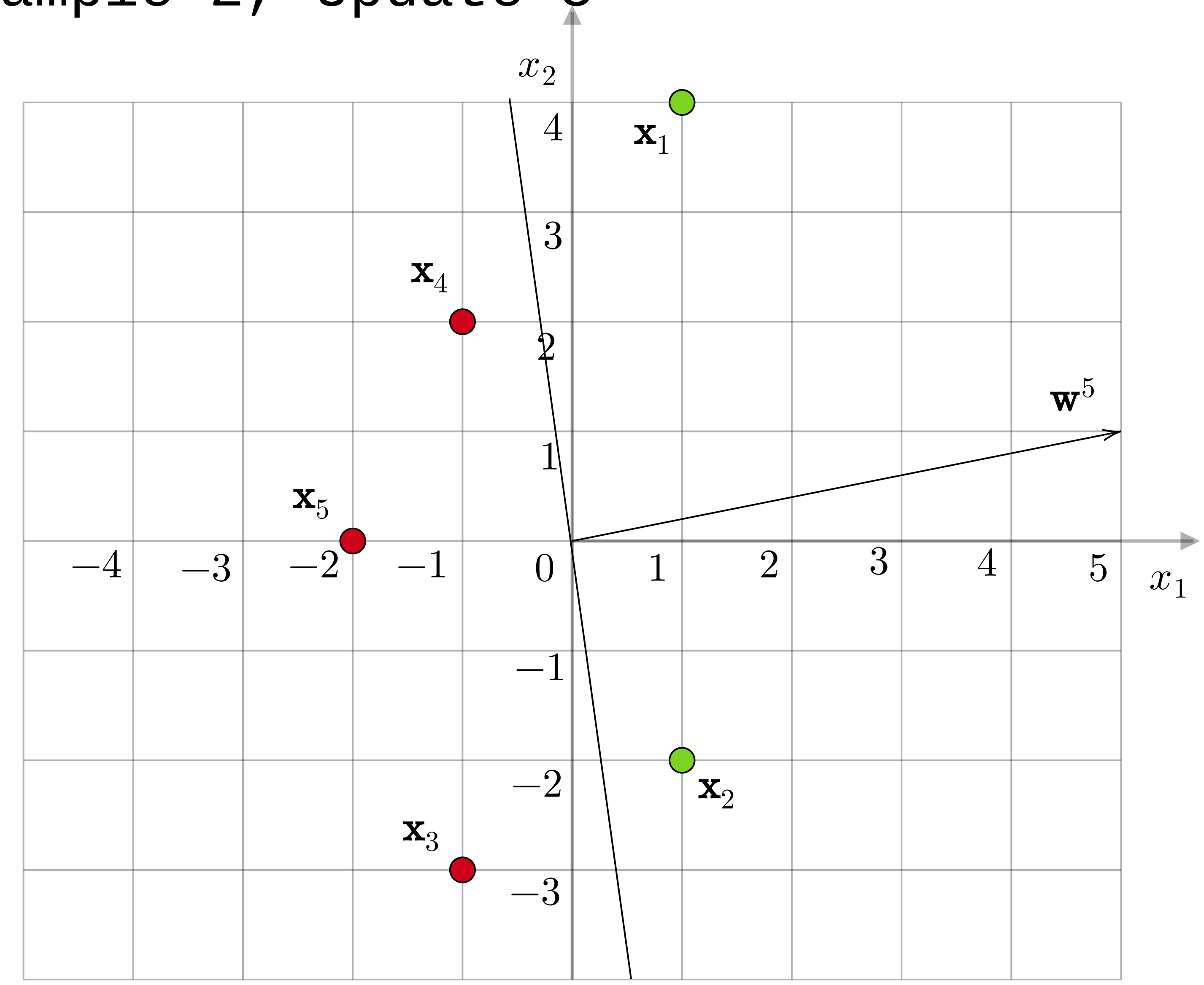
$$\begin{aligned} \mathbf{w}^4 &= \mathbf{w}^3 + \mathbf{x}_1 y_1 \\ &= \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ 3 \end{bmatrix} \end{aligned}$$

Example-2, Update-5



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

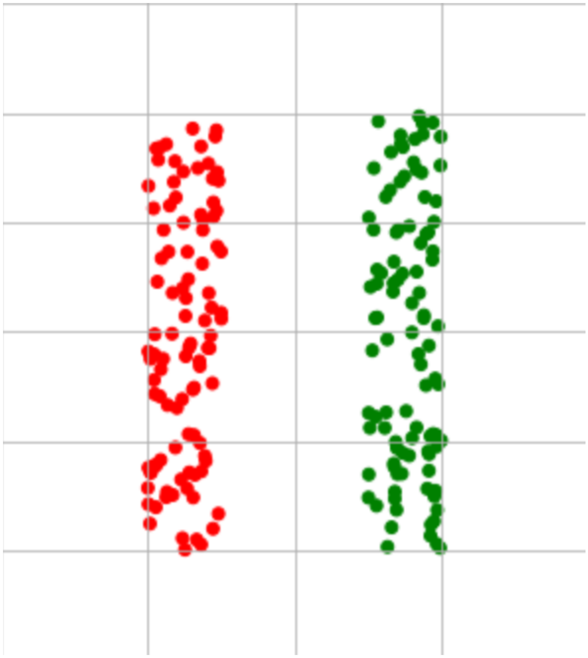
Example-2, Update-5



$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -2 \\ 4 & -2 & -3 & 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

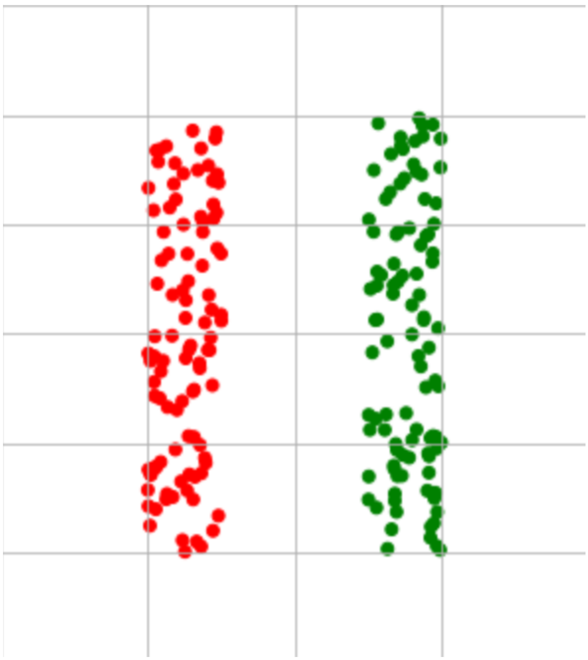
$$\begin{aligned} \mathbf{w}^5 &= \mathbf{w}^4 + \mathbf{x}_2 y_2 \\ &= \begin{bmatrix} 4 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 1 \end{bmatrix} \end{aligned}$$

Effect of Margin

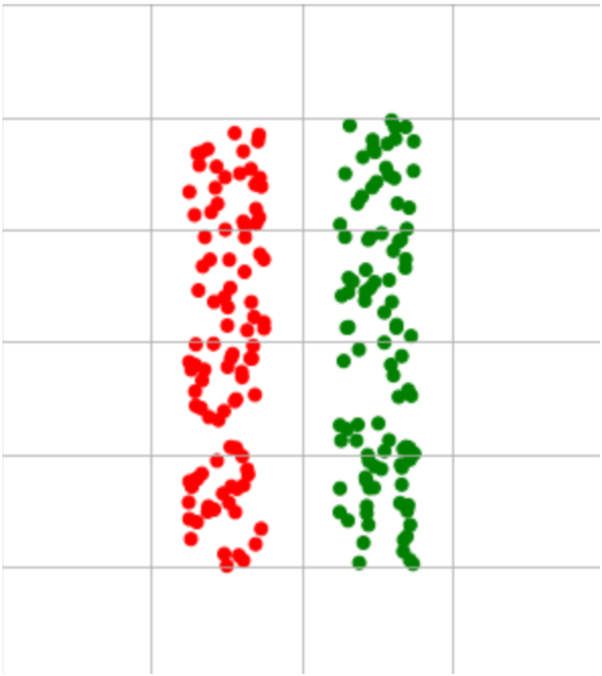


$d = 1$
$t = 1$

Effect of Margin

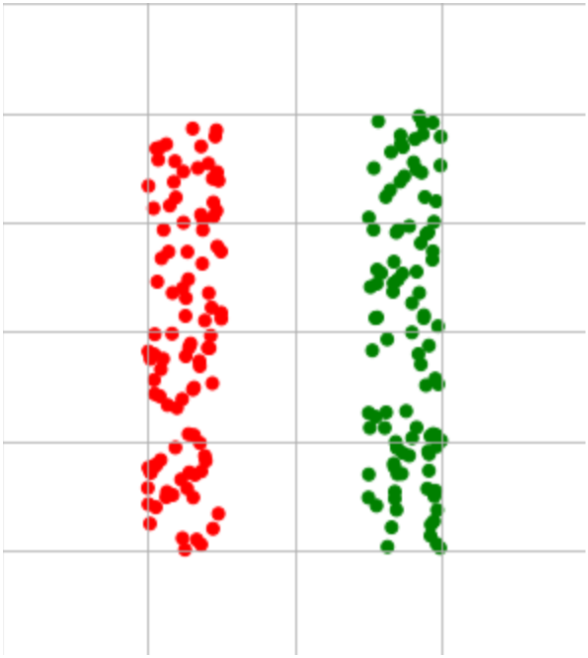


$d = 1$
$t = 1$

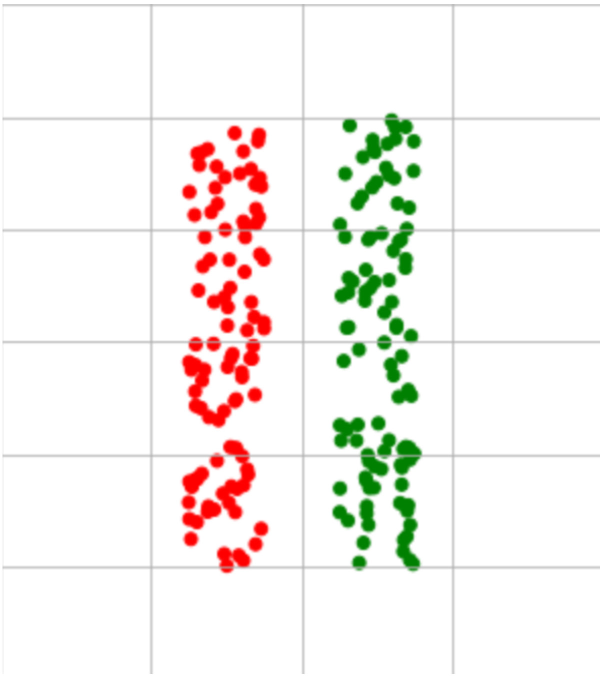


$d = 0.5$
$t = 2$

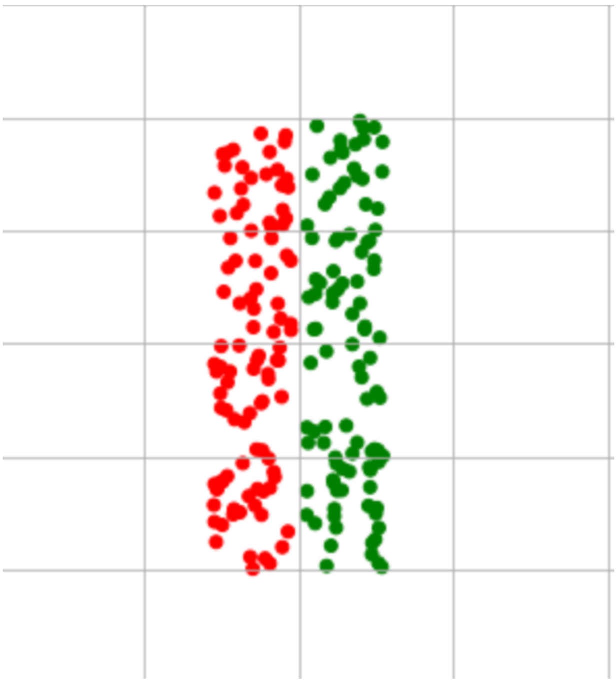
Effect of Margin



$d = 1$
$t = 1$

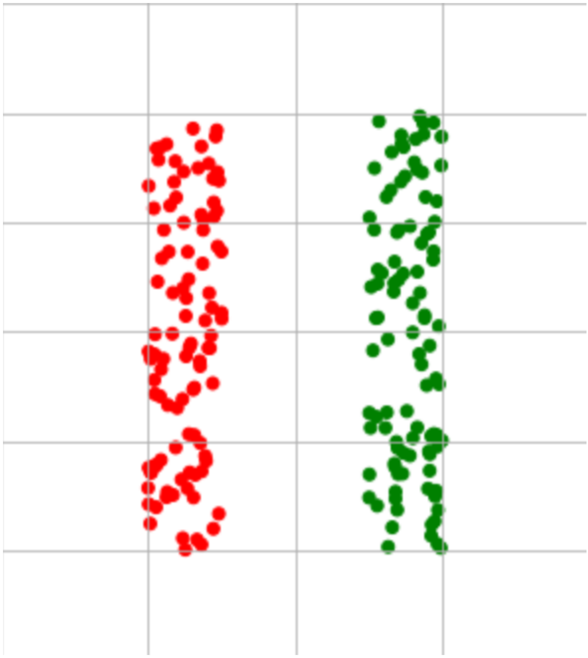


$d = 0.5$
$t = 2$

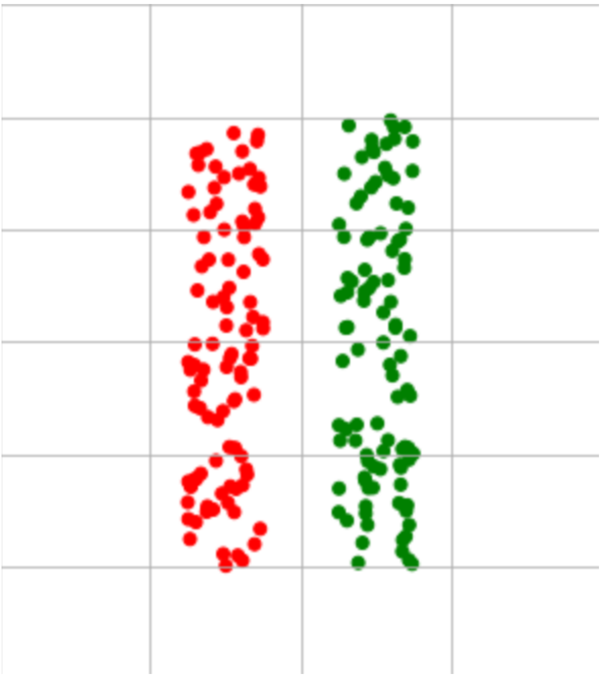


$d = 0.1$
$t = 7$

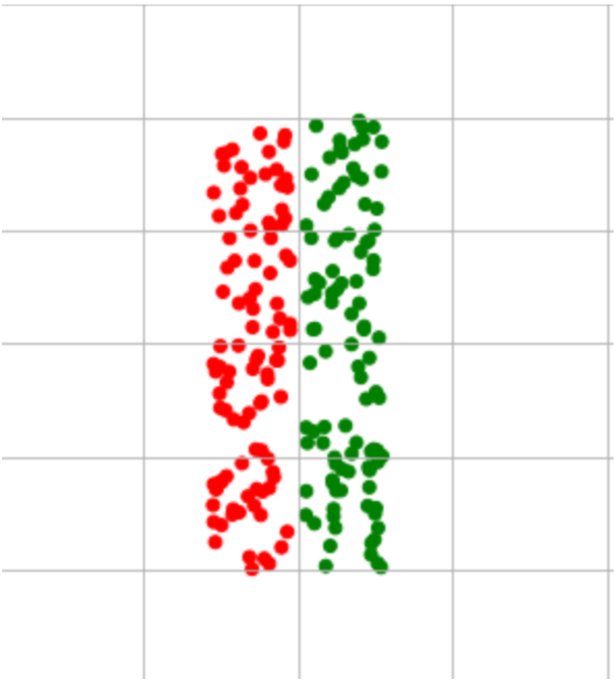
Effect of Margin



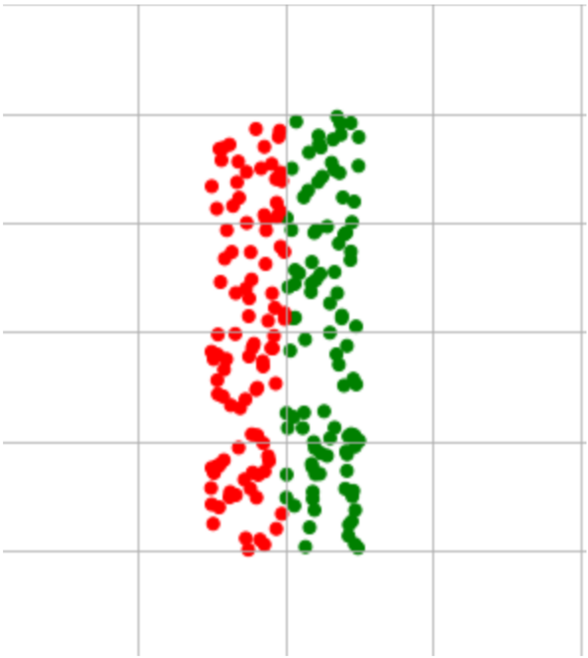
$d = 1$
$t = 1$



$d = 0.5$
$t = 2$



$d = 0.1$
$t = 7$



$d = 0.01$
$t = 44$