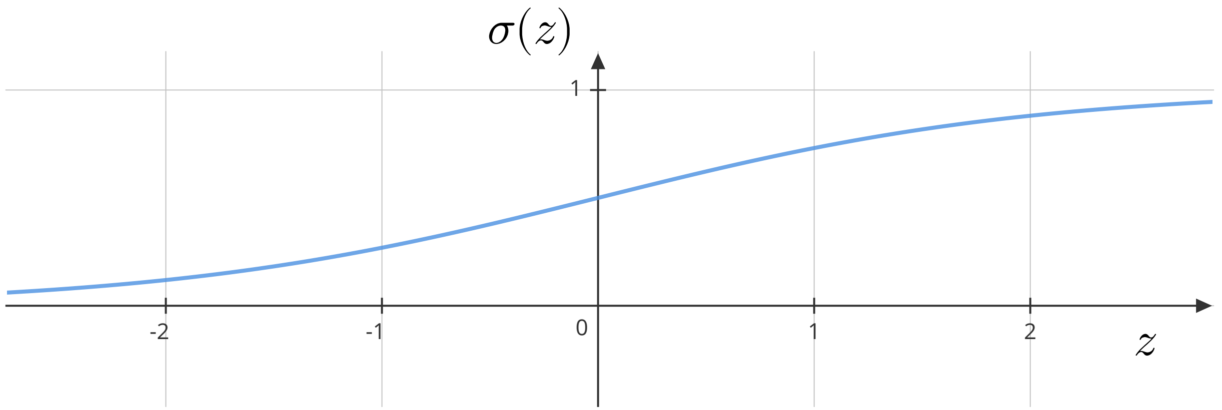
Logistic Regression

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.
- Machine Learning Refined, Second Edition
- The method of incrementally displaying content on slides is borrowed from professor Mitesh Khapra.
- These slides were prepared using the tool mathcha.io.

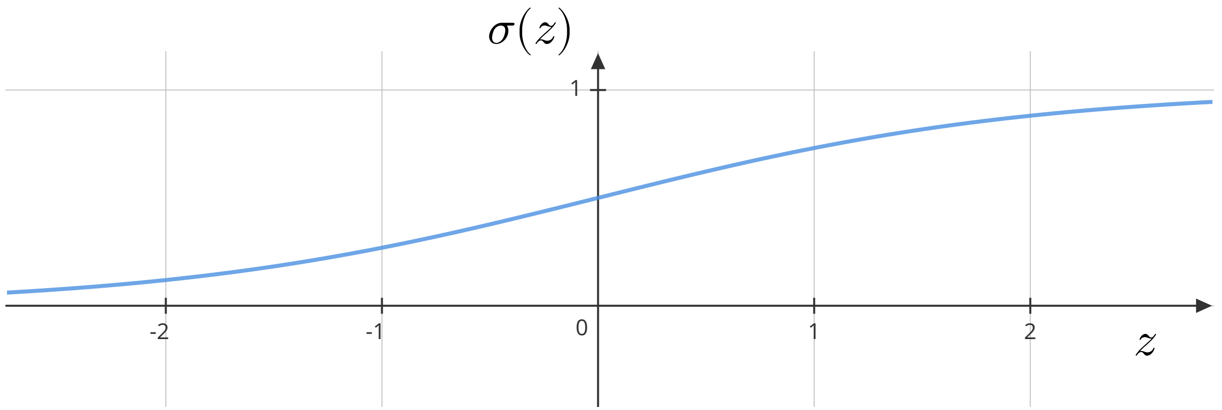
Sigmoid Function



$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid Function

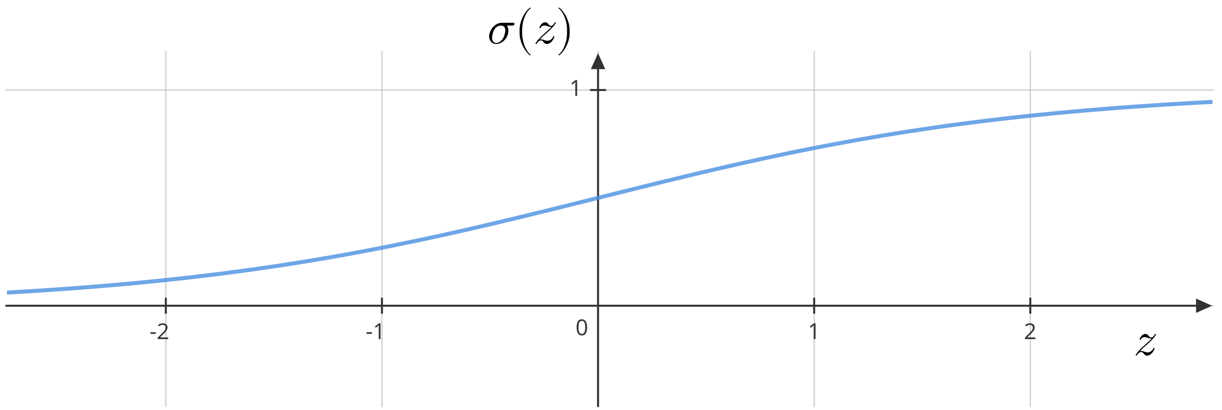
$$\sigma(z) = \begin{cases} \geqslant 0.5, & z \geqslant 0 \\ < 0.5 & z < 0 \end{cases}$$



$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid Function

$$\sigma(z) = \begin{cases} \geqslant 0.5, & z \geqslant 0 \\ < 0.5 & z < 0 \end{cases}$$



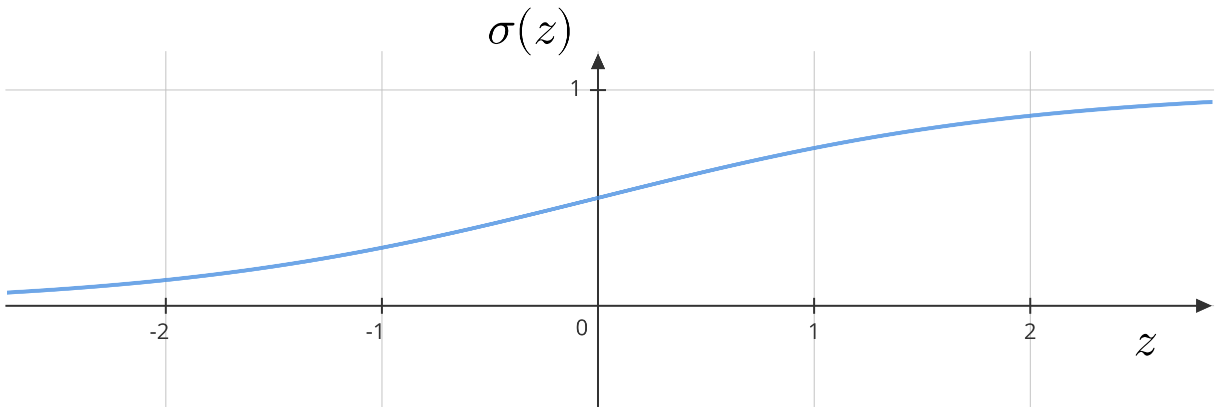
$$\lim_{z \to \infty} \sigma(z) = 1 \qquad \lim_{z \to -\infty} \sigma(z) = 0$$

$$\sigma'(z) =$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid Function

$$\sigma(z) = \begin{cases} \geqslant 0.5, & z \geqslant 0 \\ < 0.5 & z < 0 \end{cases}$$



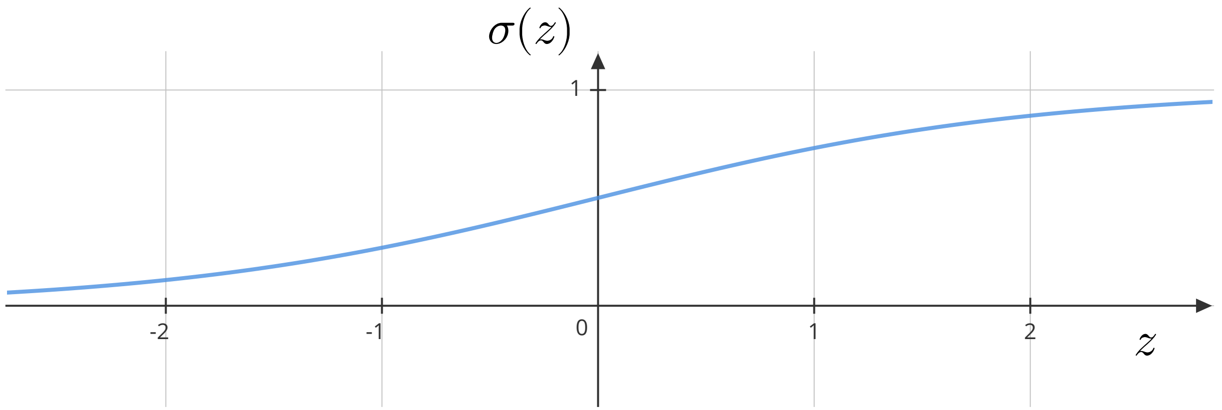
$$\lim_{z\to\infty} \sigma(z) = 1 \qquad \lim_{z\to-\infty} \sigma(z) = 0$$

$$\sigma'(z) = \frac{-1}{\left(1+e^{-z}\right)^2} \cdot e^{-z} \cdot (-1)$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Sigmoid Function

$$\sigma(z) = \begin{cases} \geqslant 0.5, & z \geqslant 0 \\ < 0.5 & z < 0 \end{cases}$$



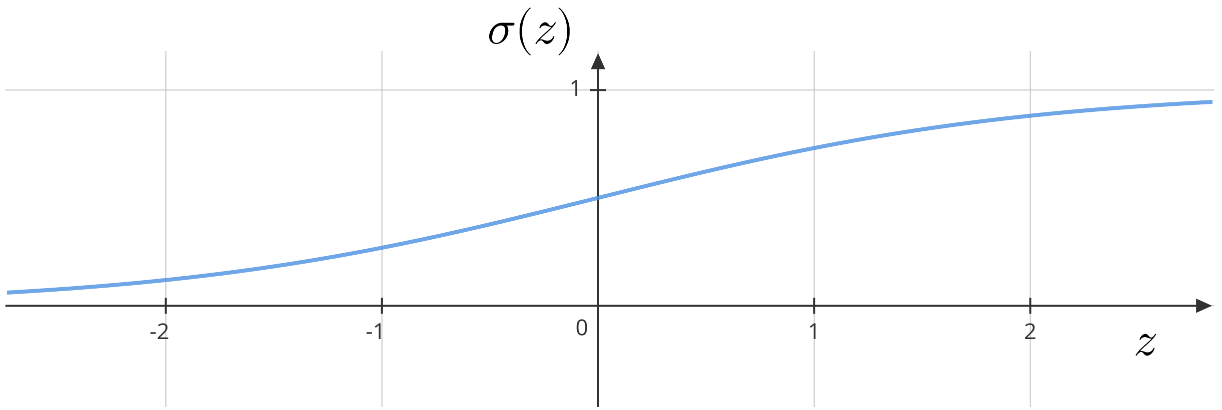
$$\lim_{z \to \infty} \sigma(z) = 1 \qquad \lim_{z \to -\infty} \sigma(z) = 0$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\begin{aligned} \sigma'(z) &= \frac{-1}{\left(1+e^{-z}\right)^2} \cdot e^{-z} \cdot (-1) \\ &= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} \end{aligned}$$

Sigmoid Function

$$\sigma(z) = \begin{cases} \geqslant 0.5, & z \geqslant 0 \\ < 0.5 & z < 0 \end{cases}$$



$$\lim_{z \to \infty} \sigma(z) = 1 \qquad \lim_{z \to -\infty} \sigma(z) = 0$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\begin{aligned} \sigma'(z) &= \frac{-1}{\left(1 + e^{-z}\right)^2} \cdot e^{-z} \cdot (-1) \\ &= \frac{1}{1 + e^{-z}} \cdot \frac{e^{-z}}{1 + e^{-z}} \\ &= \sigma(z)[1 - \sigma(z)] \end{aligned}$$

Logistic Regression

Model

$$P(y = 1 \mid \mathbf{x}) = \sigma\big(\mathbf{w}^T \mathbf{x}\big)$$

Discriminative

$$h : \mathbb{R}^d \to \{0, 1\}$$

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

- T is a threshold
- Typically, $T = 0.5$

Logistic Regression

MLE

$$\sigma_i = \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)$$

σ_i is the probability of seeing label 1

$$Y_i \mid \mathbf{x}_i \sim Br(\sigma_i)$$

Likelihood

$$\begin{aligned} L(\mathbf{w}; D) &= P(D \mid \mathbf{w}) \\ &= \prod_{i=1}^{n} \sigma_i^{y_i} \cdot (1 - \sigma_i)^{1-y_i} \end{aligned}$$

$$\max_{\mathbf{w}} \quad l(\mathbf{w}; D)$$

Log-Likelihood

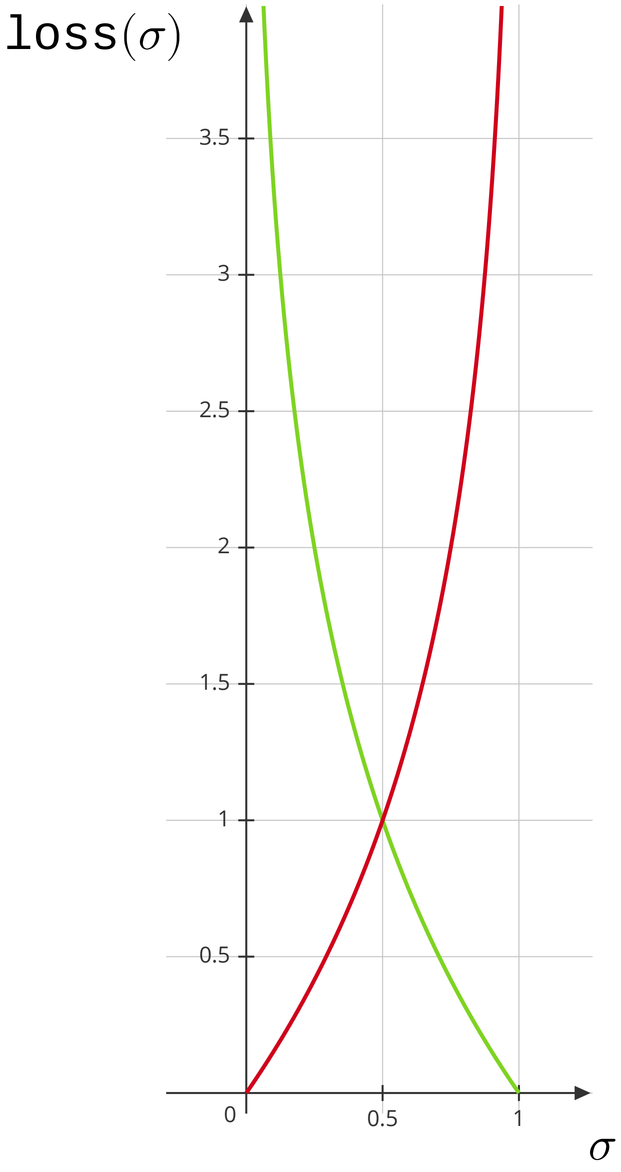
$$\begin{aligned} l(\mathbf{w}; D) &= \log L(\mathbf{w}; D) \\ &= \sum_{i=1}^{n} y_i \log \sigma_i + (1 - y_i)\log(1 - \sigma_i) \\ &= \sum_{i=1}^{n} y_i \log \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) + (1 - y_i)\log\left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \end{aligned}$$

Logistic Regression

Loss

$$\texttt{loss}(\mathbf{w}) = \sum_{i=1}^{n} -y_i \log \sigma_i - (1-y_i)\log(1-\sigma_i)$$

negative log-likelihood



$$\texttt{loss}(\sigma) = \begin{cases} -\log \sigma, & y = 1 \\ -\log(1-\sigma), & y = 0 \end{cases}$$

$$-y \log \sigma - (1-y)\log(1-\sigma)$$

cross-entropy

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big) + (1 - y_i) \log\Big[1 - \sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big)\Big]$$

Optimization

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \big[-\sigma'\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \cdot \mathbf{x}_i$$

Logistic Regression Optimization

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big) + (1 - y_i) \log\Big[1 - \sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big)\Big]$$

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big)} \cdot \sigma'\Big(\mathbf{w}^T \mathbf{x}_i\Big) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\Big(\mathbf{w}^T \mathbf{x}_i\Big)} \cdot \Big[-\sigma'\Big(\mathbf{w}^T \mathbf{x}_i\Big)\Big] \cdot \mathbf{x}_i$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) + (1 - y_i) \log\left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) + (1 - y_i) \log\left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

Logistic Regression Optimization

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) + (1 - y_i) \log\left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right]$$

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

Logistic Regression Optimization

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \big[-\sigma'\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

$$= \sum_{i=1}^{n} y_i \mathbf{x}_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

$$= \sum_{i=1}^{n} y_i \mathbf{x}_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} \left[y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \mathbf{x}_i$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\begin{aligned}
\nabla l(\mathbf{w}; D) &= \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \big[-\sigma'\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} y_i \cdot \big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} y_i \mathbf{x}_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} \big[y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big] \mathbf{x}_i
\end{aligned}$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

$$\begin{aligned}
e_i &= y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \\
&= y_i - \sigma_i
\end{aligned}$$

Logistic Regression

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

Optimization

$$\begin{aligned}
\nabla l(\mathbf{w}; D) &= \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} y_i \mathbf{x}_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i \\
&= \sum_{i=1}^{n} \left[y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \mathbf{x}_i \\
&= \sum_{i=1}^{n} e_i \cdot \mathbf{x}_i
\end{aligned}$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

$$\begin{aligned}
e_i &= y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \\
&= y_i - \sigma_i
\end{aligned}$$

Logistic Regression Optimization

$$l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \log \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) + (1 - y_i) \log\big[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\big]$$

$$\nabla l(\mathbf{w}; D) = \sum_{i=1}^{n} y_i \cdot \frac{1}{\sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \sigma'\left(\mathbf{w}^T \mathbf{x}_i\right) \cdot \mathbf{x}_i + (1 - y_i) \cdot \frac{1}{1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)} \cdot \left[-\sigma'\left(\mathbf{w}^T \mathbf{x}_i\right)\right] \cdot \mathbf{x}_i$$

$$= \sum_{i=1}^{n} y_i \cdot \left[1 - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \cdot \mathbf{x}_i - (1 - y_i) \cdot \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

$$\sigma'(z) = \sigma(z)[1 - \sigma(z)]$$

$$= \sum_{i=1}^{n} y_i \mathbf{x}_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \cdot \mathbf{x}_i$$

$$\begin{aligned} e_i &= y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big) \\ &= y_i - \sigma_i \end{aligned}$$

$$= \sum_{i=1}^{n} \left[y_i - \sigma\big(\mathbf{w}^T \mathbf{x}_i\big)\right] \mathbf{x}_i$$

$$= \sum_{i=1}^{n} e_i \cdot \mathbf{x}_i$$

$$\boxed{\texttt{gradient} = \texttt{error} \times \texttt{data-point}}$$

LogisticRegression(D):

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

    converged = False

    while not converged:

        $\mathbf{w}^{t+1} = \mathbf{w}^t + \alpha \sum_{i=1}^{n} \Big(y_i - \sigma\big({\mathbf{w}^t}^T \mathbf{x}_i\big)\Big)\mathbf{x}_i$

        converged = check(criterion)

    return $\mathbf{w}^*$

Gradient Ascent

LogisticRegression(D):

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

$\quad$ converged = False

$\quad$ while not converged:

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

$$\mathbf{w}^{t+1} = \mathbf{w}^t + \alpha\Big(y_i - \sigma\big({\mathbf{w}^t}^T \mathbf{x}_i\big)\Big)\mathbf{x}_i$$

$\qquad$ converged = check(criterion)

$\quad$ return $\mathbf{w}^*$

Stochastic Gradient Ascent

Logistic Regression vs Perceptron

LogisticRegression(D) :

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

    converged = False

    while not converged :

        for $i = 1$ to n :

            $\boxed{\mathbf{w}^{t+1} = \mathbf{w}^t + \alpha\Big(y_i - \sigma\Big({\mathbf{w}^t}^T \mathbf{x}_i\Big)\Big)\mathbf{x}_i}$

        converged = check(criterion)

    return $\mathbf{w}^*$

Logistic Regression vs Perceptron

LogisticRegression(D):

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

 converged = False

 while not converged:

  for $i = 1$ to n:

$$\boxed{\mathbf{w}^{t+1} = \mathbf{w}^t + \alpha\Big(y_i - \sigma\Big({\mathbf{w}^t}^T\mathbf{x}_i\Big)\Big)\mathbf{x}_i}$$

  converged = check(criterion)

 return $\mathbf{w}^*$

Perceptron(D):

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

 converged = False

 while not converged:

  for $i = 1$ to n:

$$\boxed{\mathbf{w}^{t+1} = \mathbf{w}^t + \frac{1}{2}\Big(y_i - \mathbf{1}\Big[{\mathbf{w}^t}^T\mathbf{x}_i \geqslant 0\Big]\Big)\mathbf{x}_i}$$

  converged = check(criterion)

 return $\mathbf{w}^*$

Logistic Regression vs Perceptron

LogisticRegression(D):

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

    converged = False

    while not converged:

        for $i = 1$ to n:

            $\boxed{\mathbf{w}^{t+1} = \mathbf{w}^t + \alpha\Big(y_i - \sigma\Big({\mathbf{w}^t}^T\mathbf{x}_i\Big)\Big)\mathbf{x}_i}$

        converged = check(criterion)

    return $\mathbf{w}^*$

Soft Prediction

Perceptron(D):

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

    converged = False

    while not converged:

        for $i = 1$ to n:

            $\boxed{\mathbf{w}^{t+1} = \mathbf{w}^t + \frac{1}{2}\Big(y_i - \mathbf{1}\Big[{\mathbf{w}^t}^T\mathbf{x}_i \geqslant 0\Big]\Big)\mathbf{x}_i}$

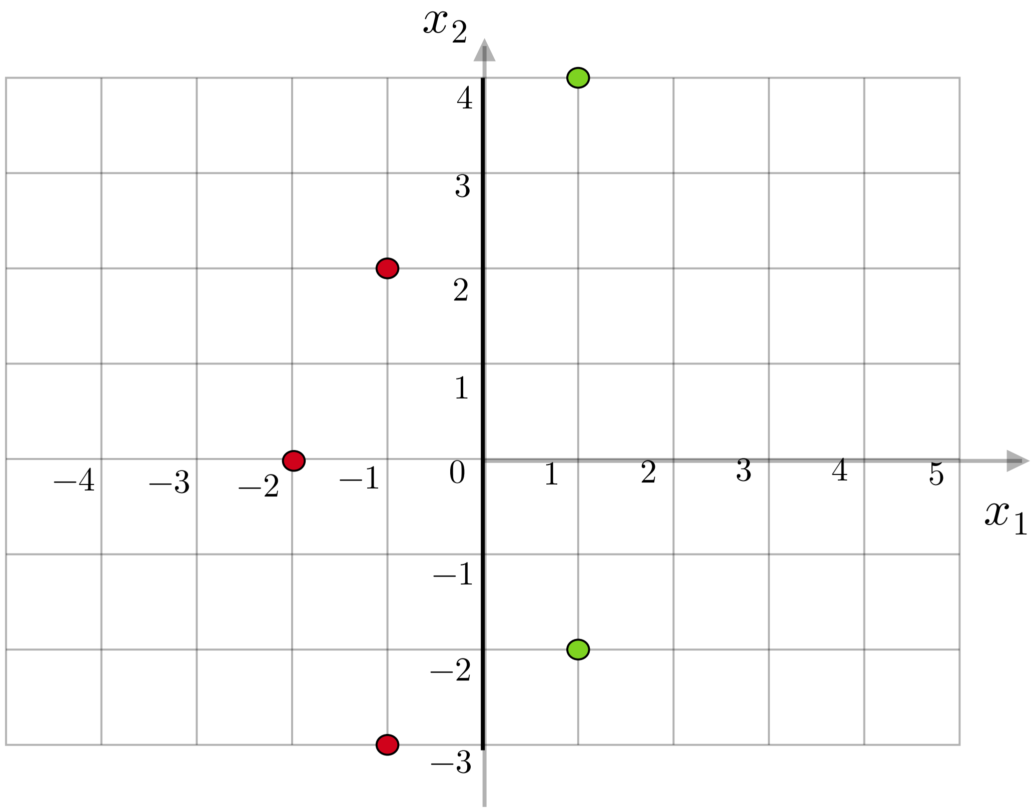
        converged = check(criterion)

    return $\mathbf{w}^*$

Hard Prediction

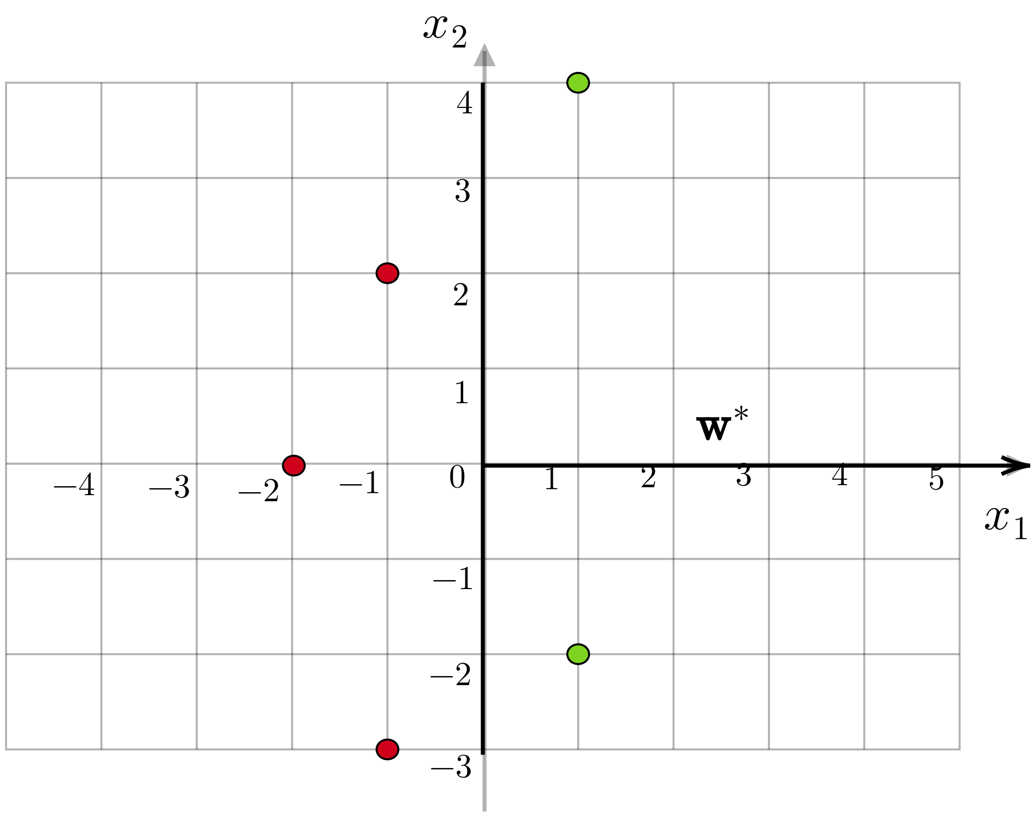
Example-1

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

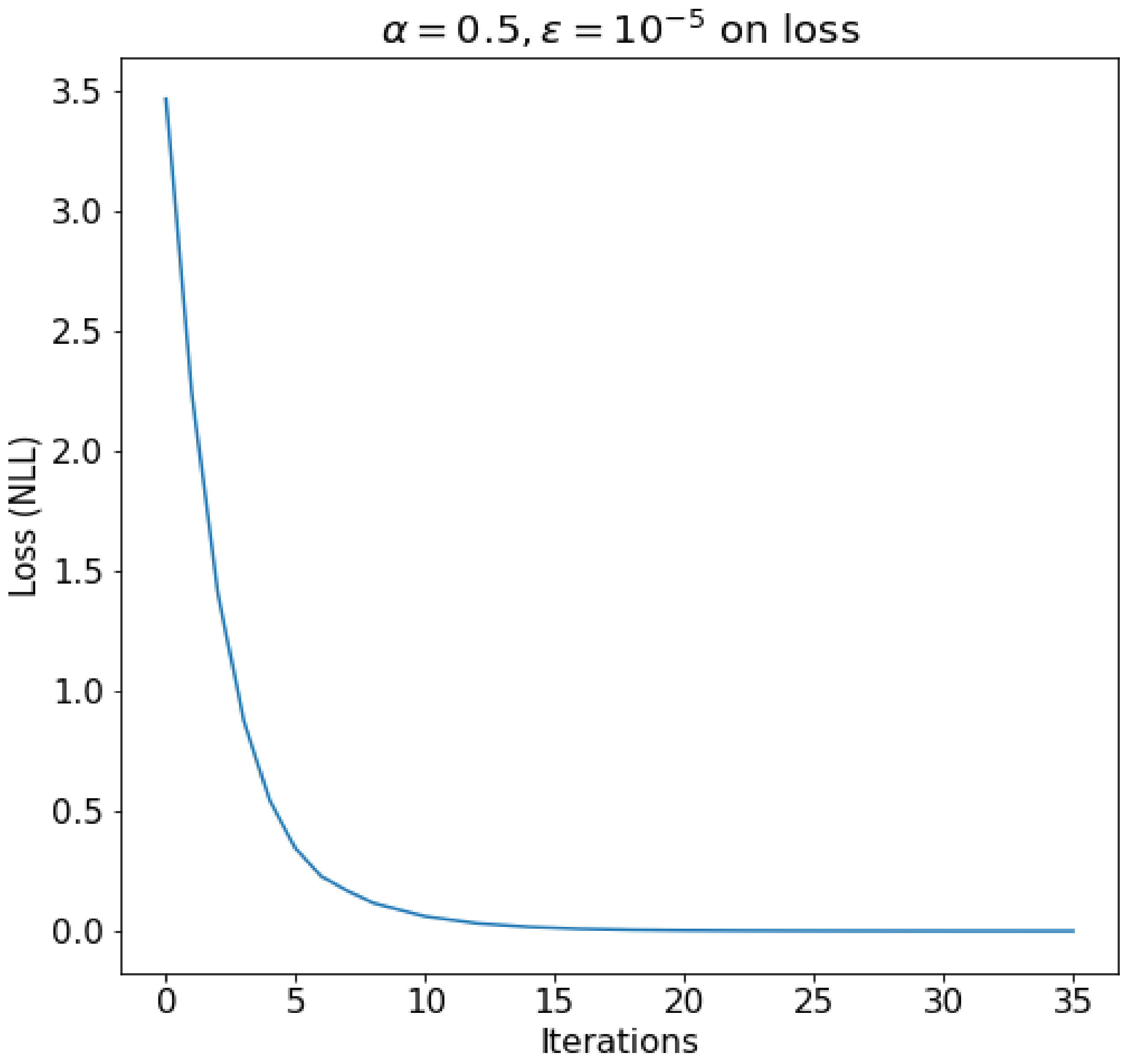


Example-1

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



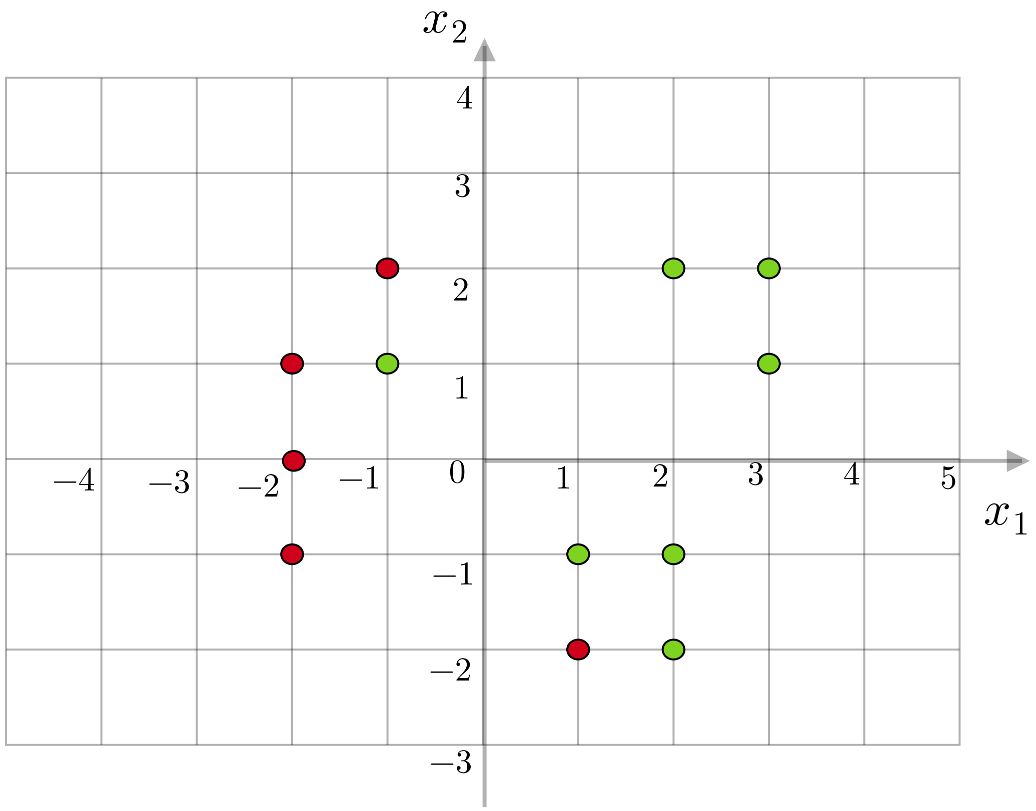
$$\mathbf{w}^* = \begin{bmatrix} 12.5 \\ -0.06 \end{bmatrix}$$



Example-2

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

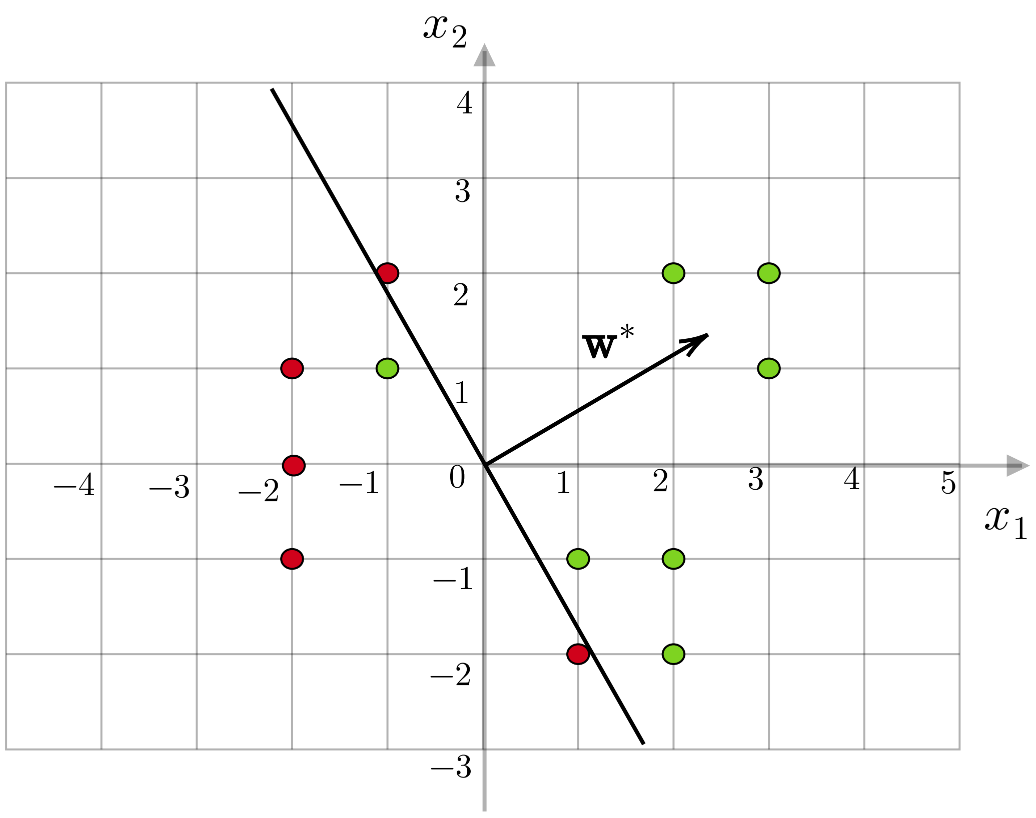
$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$



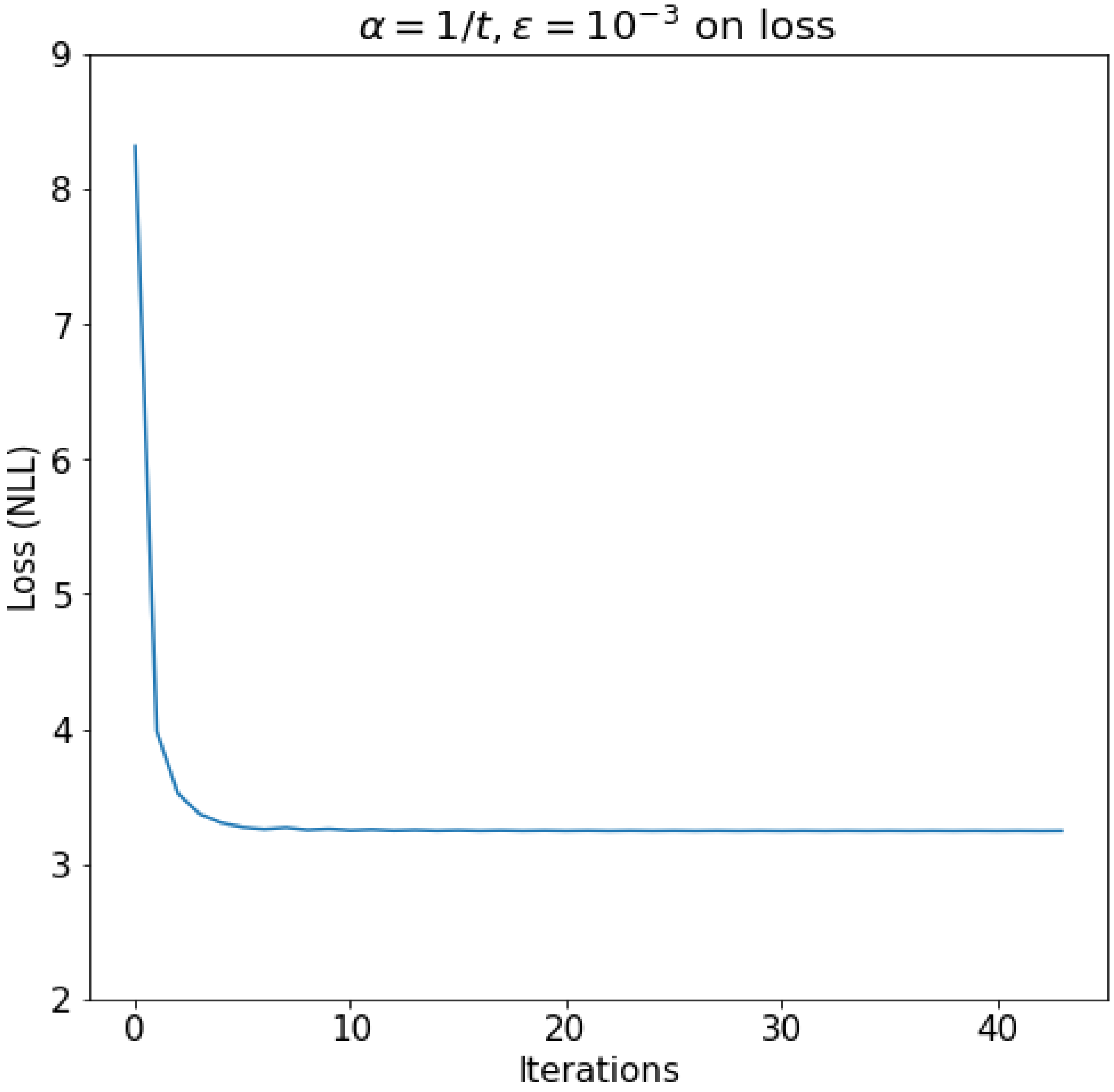
Example-2

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

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$



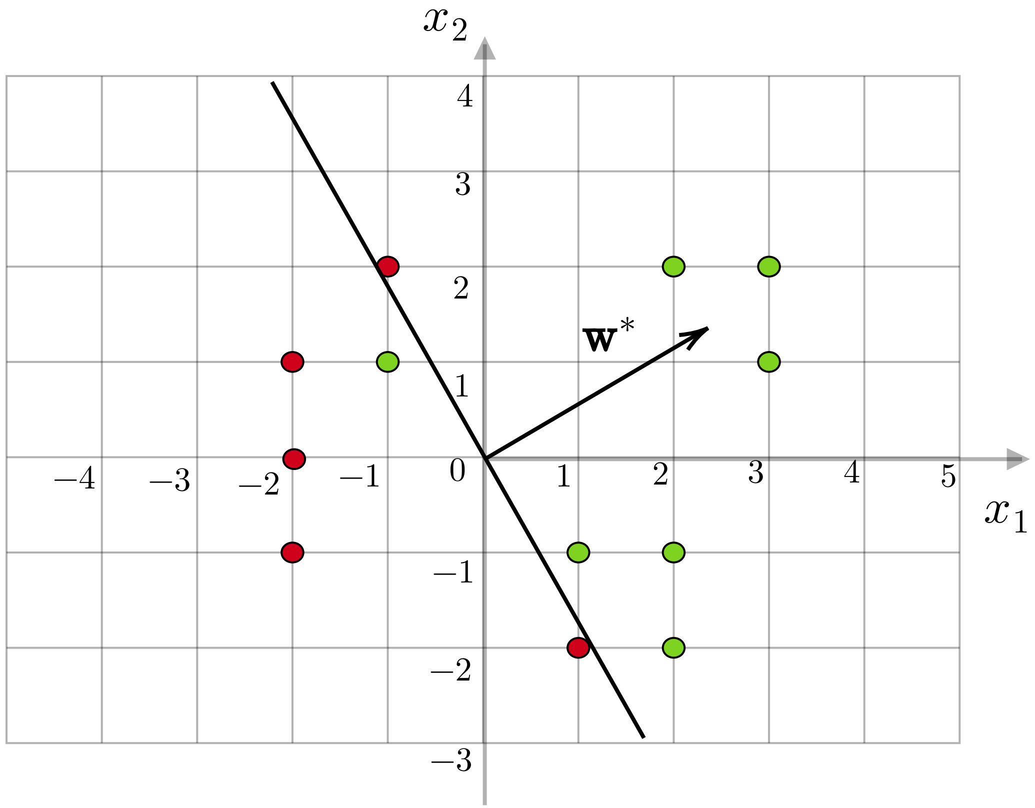
$$\mathbf{w}^* = \begin{bmatrix} 2.30 \\ 1.25 \end{bmatrix}$$



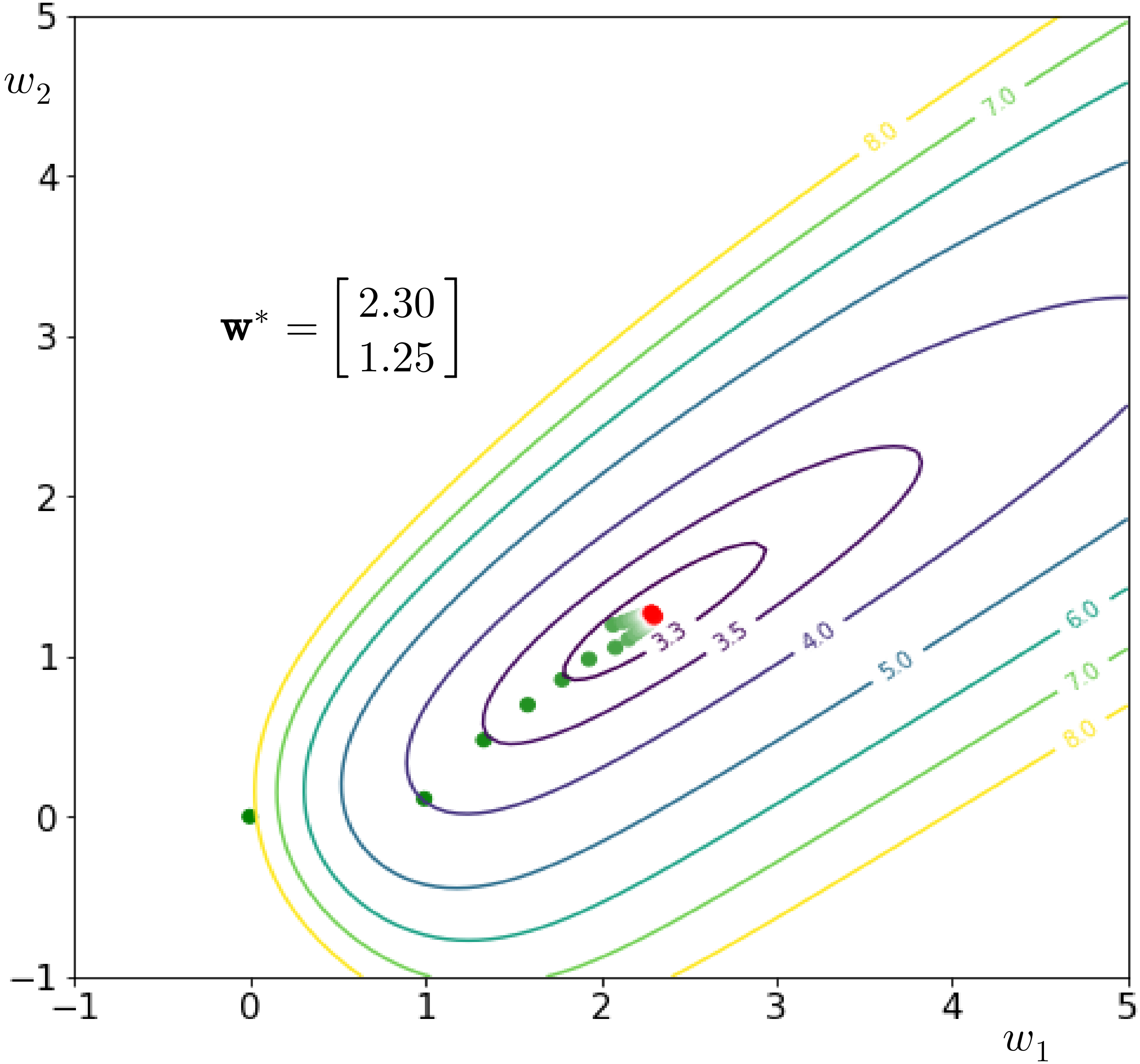
Example-2

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

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$



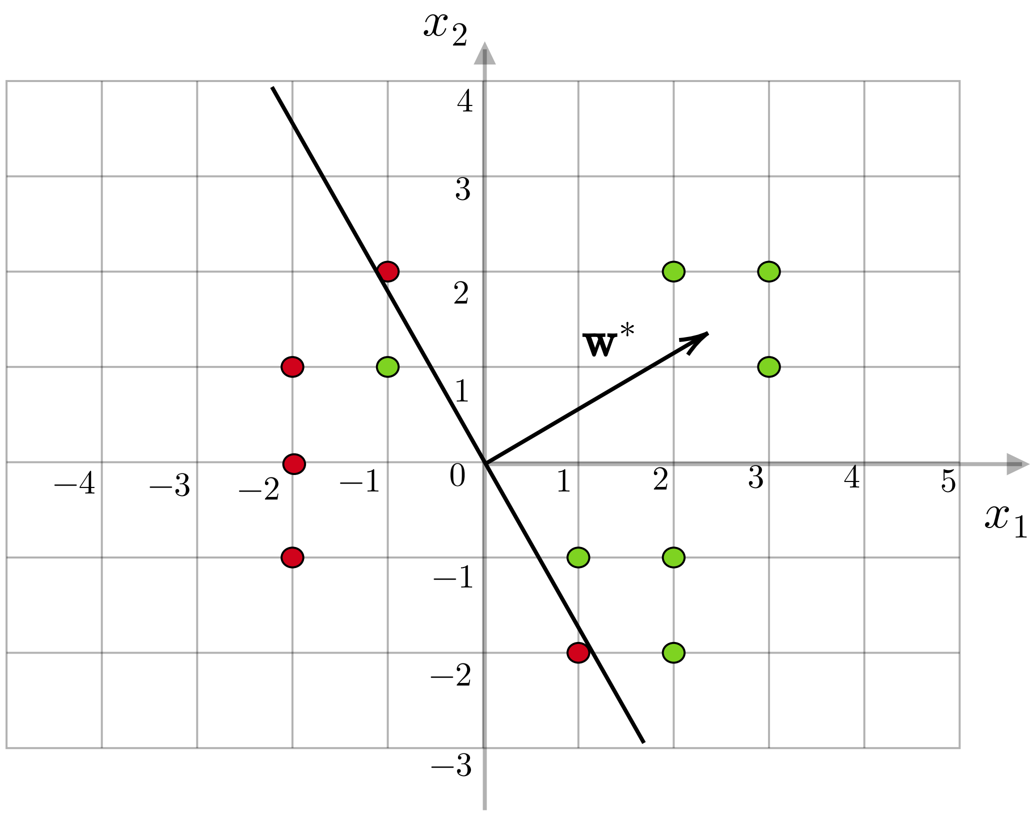
Feature Space



Weight Space

Example-2

$$\mathbf{X} = \begin{bmatrix} 2 & 3 & 3 & 1 & 2 & 2 & -1 & 1 & -1 & -2 & -2 & -2 \\ 2 & 2 & 1 & -1 & -1 & -2 & 1 & -2 & 2 & 1 & 0 & -1 \end{bmatrix},$$
$$\mathbf{y} = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$$



$$\mathbf{w}^* = \begin{bmatrix} 2.30 \\ 1.25 \end{bmatrix}$$

x_1	x_2	y	σ	$\widehat{y}$
2	2	1	0.99	1
3	2	1	0.99	1
3	1	1	0.99	1
1	−1	1	0.74	1
2	−1	1	0.96	1
2	−2	1	0.89	1
−1	1	1	0.25	0
1	−2	0	0.45	0
−1	2	0	0.54	1
−2	1	0	0.03	0
−2	0	0	0.01	0
−2	−1	0	0.002	0