Optimization in Machine Learning

Machine Learning Academy

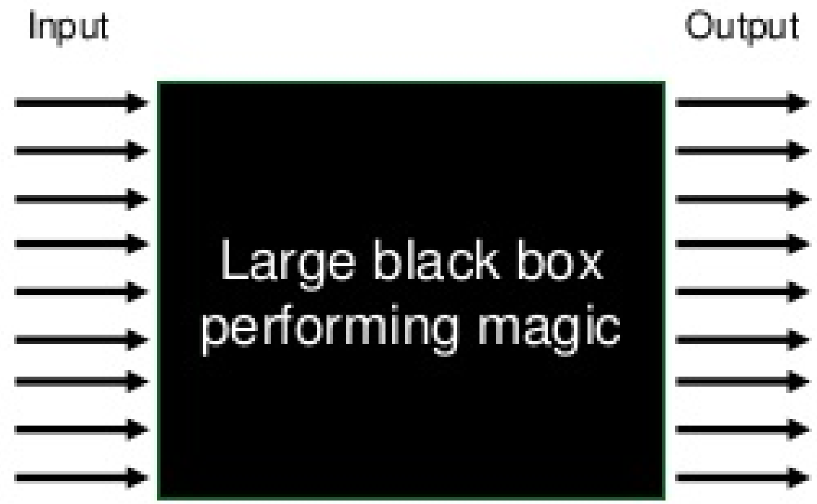
Outline

1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Outline

1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Supervised Learning (1)



$$y = f(x, \theta)$$

Supervised Learning (2)

But how do we find these parameters?

$$y = \mathrm{f}(x, \theta)$$

How do we know what a good parameter is?

$$\hat{y} = \mathrm{f}(x, \theta)$$

Minimize the prediction error! $\longrightarrow$ $e = |y - \hat{y}|$

Outline

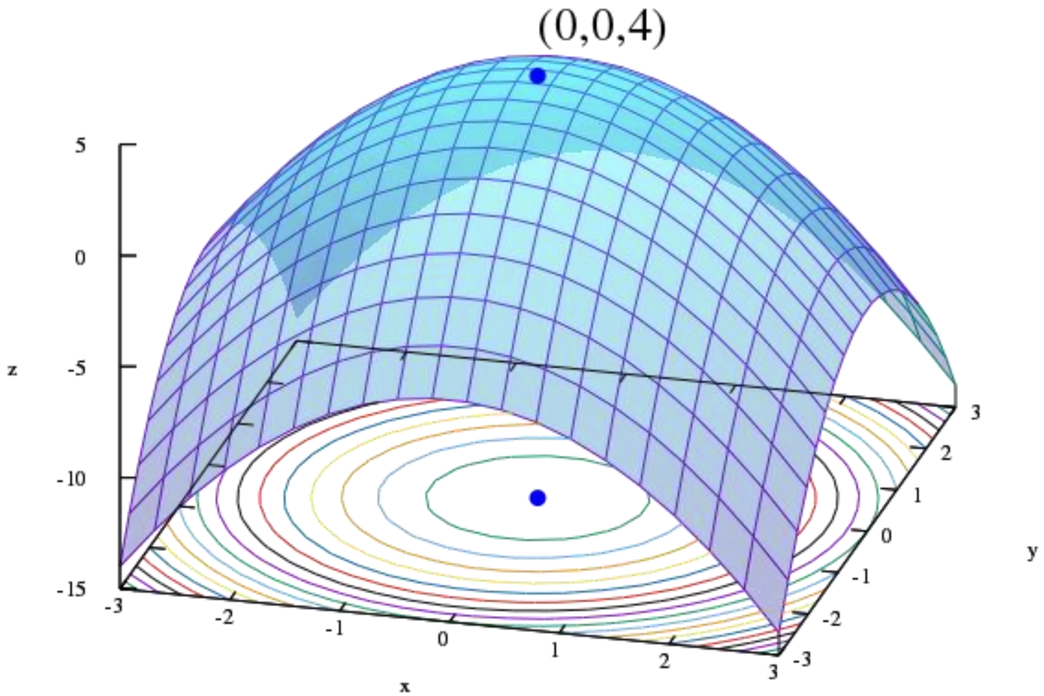
1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

How does optimization fit into this? (1)

"Optimization is the process by which we find the *best available* values of some **objective function** given a **defined domain**".

$$e = |y - \hat{y}|$$
$$= |\, y - f(x, \textcolor{red}{\theta})|$$

Changing this parameter
influences the prediction error

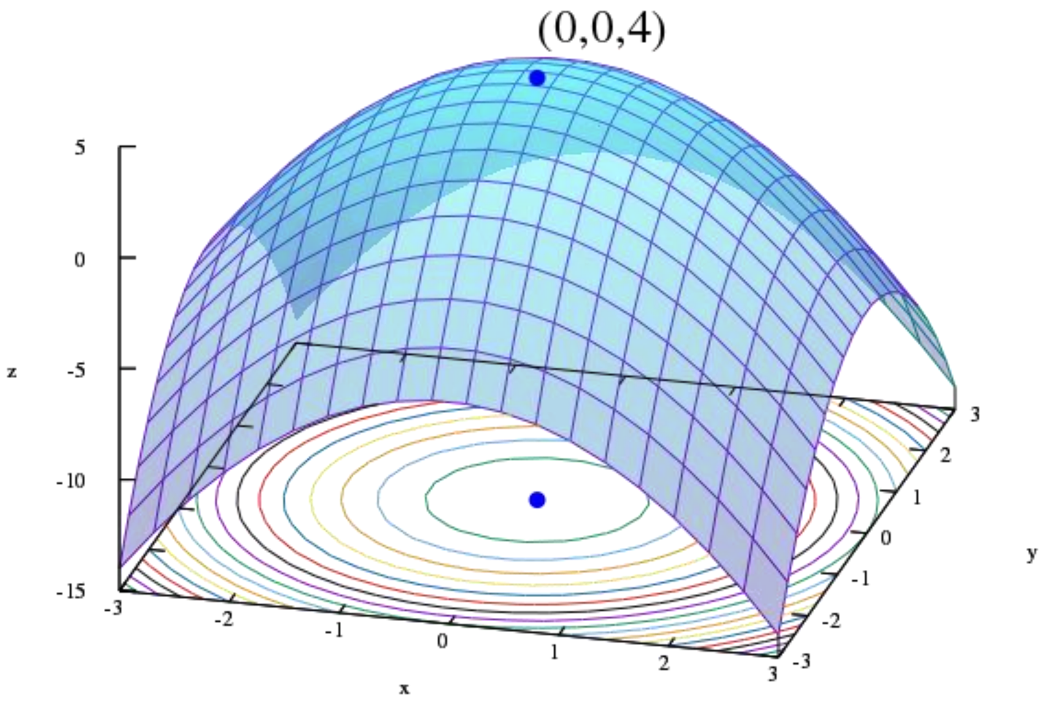


How does optimization fit into this? (2)

More formally, optimization is defined by:

$\max_\theta f(\theta)$
Subject to constraints

1. Function to minimize
2. Constraints



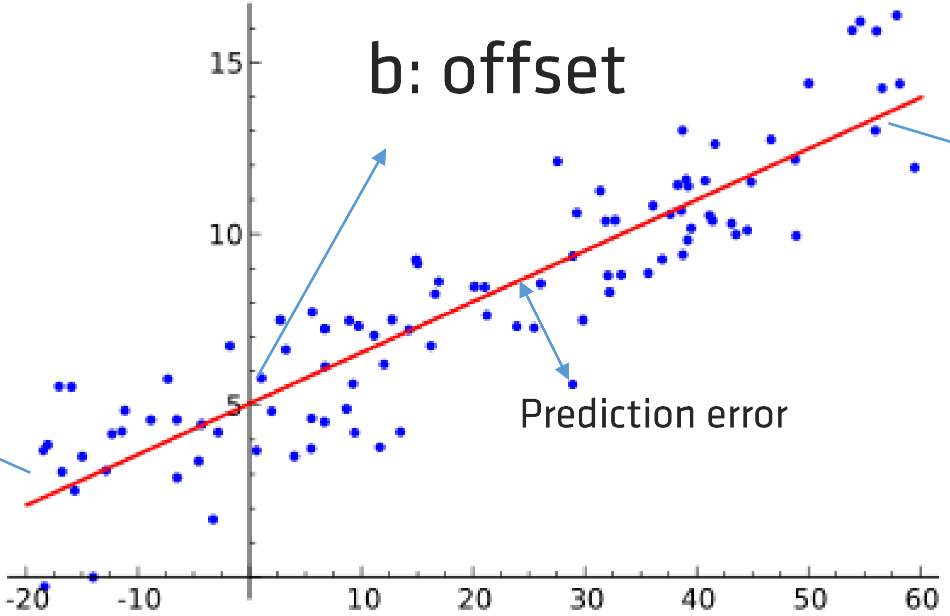
Outline

1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Linear regression (1)

$\theta = (m, b)$

$y = f(x, \theta) = mx + b$



Linear regression (2)

Application example:

$$\hat{y} = f(x, \theta) = mx + b, \quad \theta = (m, b)$$

$$\mathrm{e} = |\, y - \hat{y} \,| = |y - (mx + b)|$$

Objective

Parameters

Outline

1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Notebook 1: Introduction to CVXPy

CVXPy is a Python package to solve optimization problems

We define the problem:

1. **Objective function**
2. **Constraints**

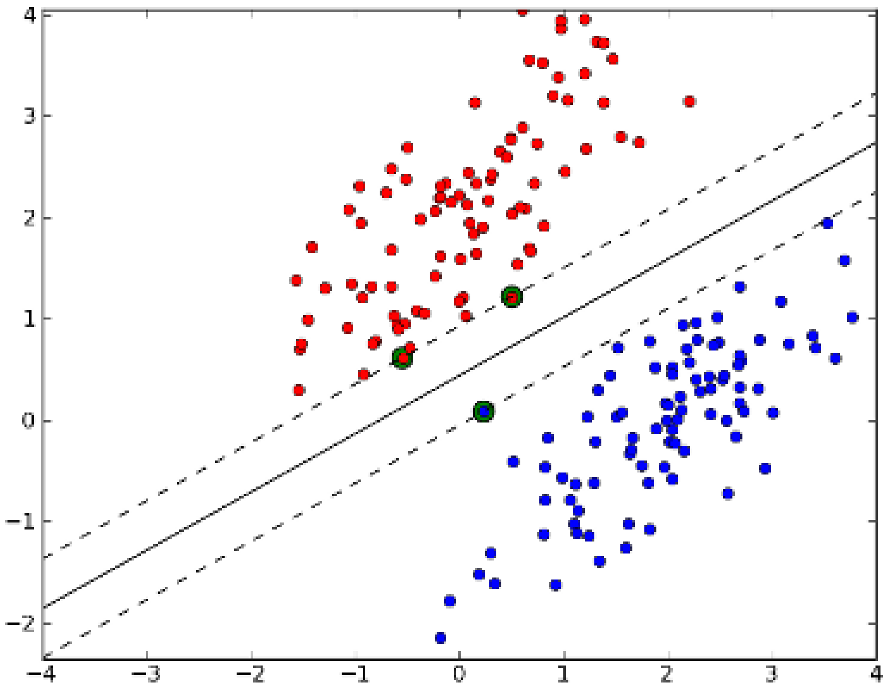
And the package solves the problem for us!

Lets look at how to define the regression problem in CVXPy!

Outline

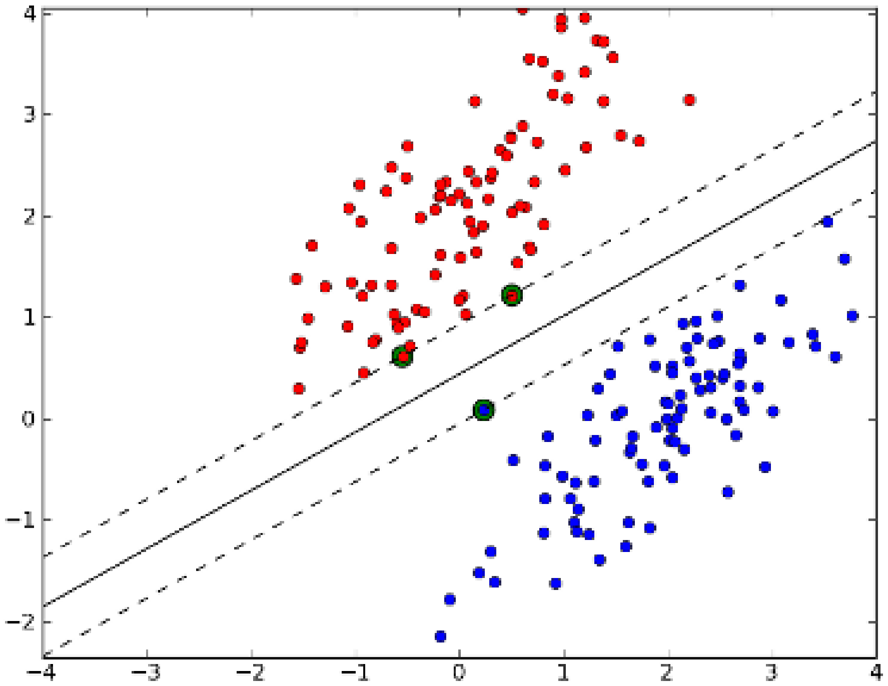
1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Support Vector Machine (1)



How to formulate this as an optimization problem?

Support Vector Machine (2)



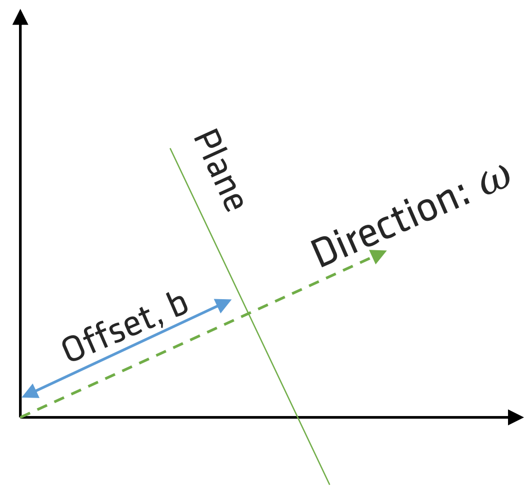
We want to find a plane which divides the two sets of data. For this we need to define:

1. Parameters of the plane: θ
2. Prediction function: f(x, θ)
3. Objective function: $e = y - f(x, \theta)$

Support Vector Machine: Parameters (3)

Defining the location of a plane:

Parameters of the plane: Direction (ω) and offset (b)



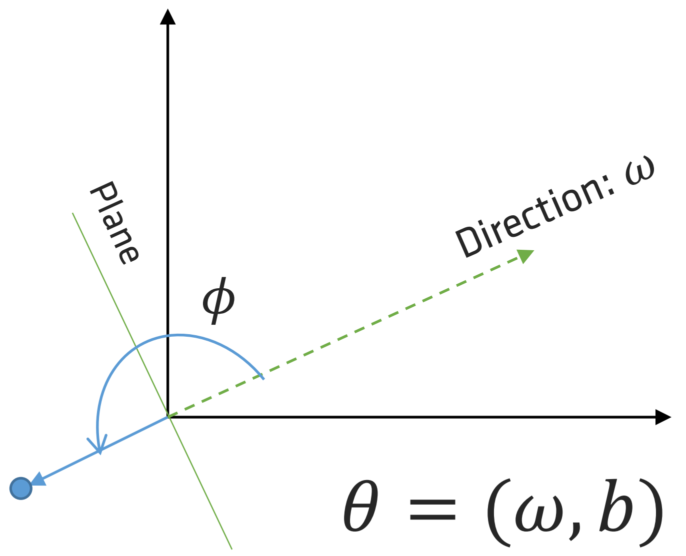
$$\theta = (\omega, b)$$

Support Vector Machine: Prediction (4)

Prediction: y = f(x, θ)

Two classes: $y_1 = -1,\ y_2 = 1,$

-1 if on the left side of the plane and 1 if on the right



$$\theta = (\omega, b)$$

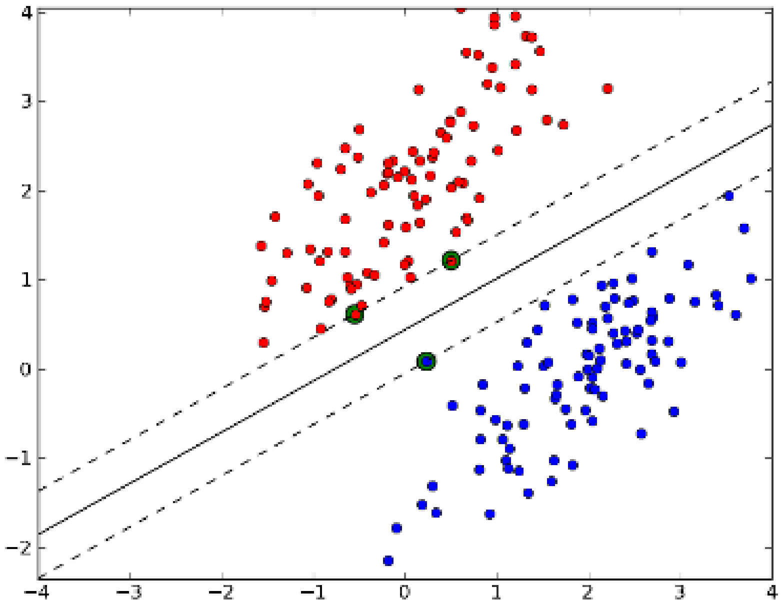
For a plane at the origin, the side of the plane is given by the dot product between the direction and the point:

$$y = \omega \cdot x = \cos(\phi)\ ||\omega||\ ||x||$$

For a plane with offset the function becomes:

$$\mathbf{y} = \boldsymbol{\omega} \cdot \boldsymbol{x} + \boldsymbol{b} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$$

Support Vector Machine: Optimization (4)



Constraints:

Find parameters θ such that all samples are correctly labelled:

$$y_i(x_i \cdot \omega + b) \geq 1$$

This gives us a lot of possible parameters, we impose a norm objective to find the smallest one:

Objective:

Minimize $||\omega||$

Outline

1. Supervised Learning
2. What is optimization?
3. Regression as an optimization problem
4. Notebook 1: Introduction to CVXPy
5. Support Vector Machine as an optimization problem
6. Notebook 2: SVM in CVXPy
7. How does this fit in with the next lecture?

Notebook 2: SVM in CVXPy

In the last few slides we defined:

1. Model parameters: $\theta = (\omega . b)$
2. Prediction function: $\mathrm{y} = f(x, \theta) = \omega \cdot x + b$
3. Objective and constraints:

$$\min_{\theta} ||\omega||, \quad s.t.\ y_i(\omega \cdot x_i + b) \geq 1$$

We will now formulate this problem in the CVXPy package and solve it!

How is this related to the next lecture?

Neural networks is an optimization problem!

In the next section we will use the same formulation to solve for the neural network parameters