Machine Learning Techniques

ML vs Traditional Programming

Dr. Ashish Tendulkar

IIT Madras

ML vs Traditional Programming

Let's take a couple of tasks and try to write programs to solve each of them:

- Task 1: Adding two numbers:

$$12 + 45 = ?$$

ML vs Traditional Programming

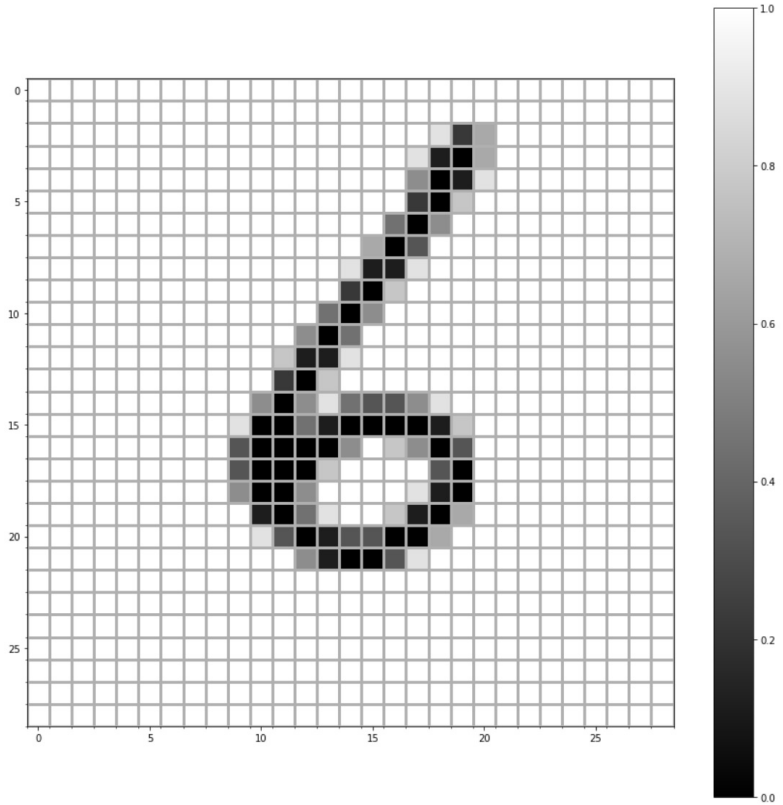
Let's take a couple of tasks and try to write programs to solve each of them:

- Task 1: Adding two numbers:

$$12 + 45 = ?$$

ML vs Traditional Programming

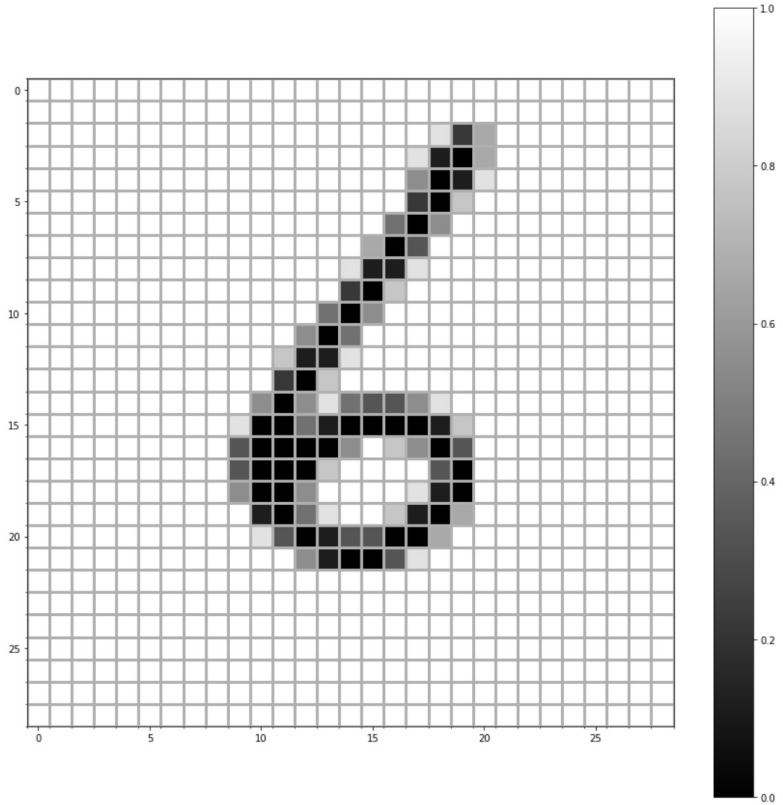
- Task 2: Recognize handwritten digit. An image is a matrix of numbers.



Which one is easier?

ML vs Traditional Programming

- Task 2: Recognize handwritten digit. An image is a matrix of numbers.



Which one is easier?

Image representation of a computer

How does a computer see an image?

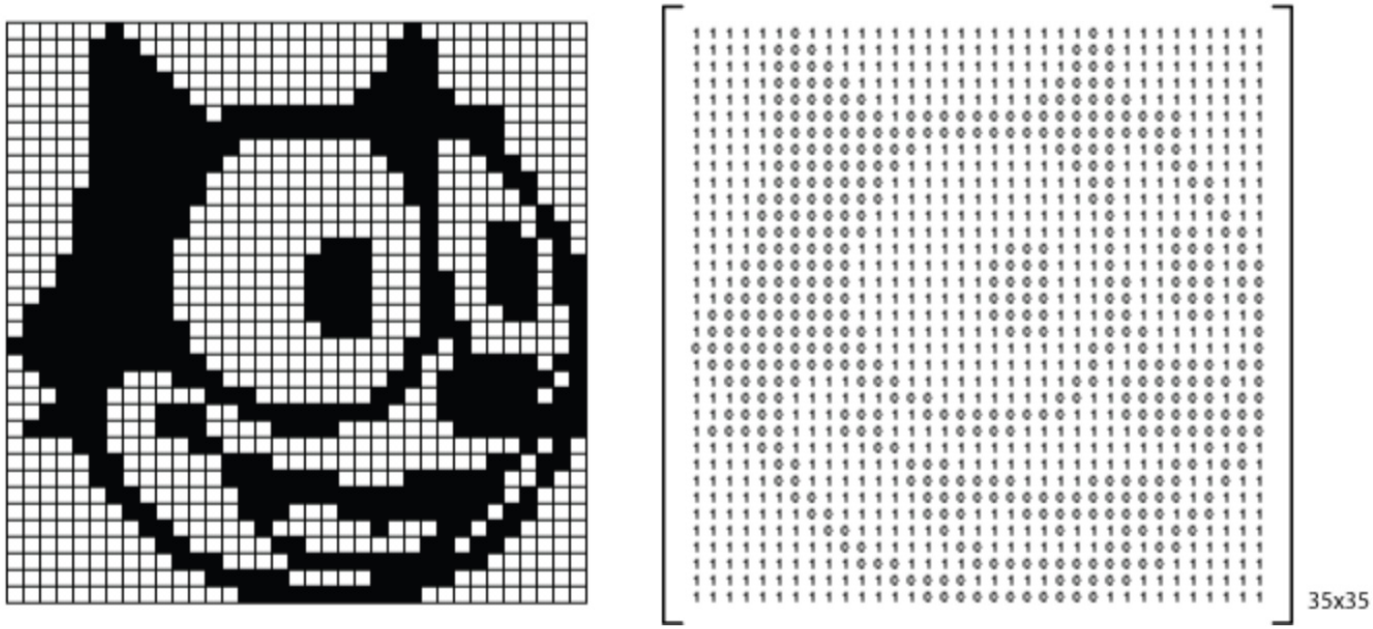


Image Source:uff.br

Adding Numbers

- Adding two numbers

$$12 + 45 = 57$$

- Trivial for any decent programmer.

Adding Numbers

- Adding two numbers

$$12 + 45 = 57$$

- Trivial for any decent programmer.

Recognise a digit

Solving the second task:

- Design a bunch of rules for solving this task, which seems like a reasonable approach.
- However, such rules are brittle and one can readily come up with examples that will break such a rule-based system.

What is a more robust way of solving this task?

Recognise a digit

Solving the second task:

- Design a bunch of rules for solving this task, which seems like a reasonable approach.
- However, such rules are brittle and one can readily come up with examples that will break such a rule-based system.

What is a more robust way of solving this task?

Recognise a digit

Solving the second task:

- Design a bunch of rules for solving this task, which seems like a reasonable approach.
- However, such rules are brittle and one can readily come up with examples that will break such a rule-based system.

What is a more robust way of solving this task?

Recognise a digit

Solving the second task:

- Design a bunch of rules for solving this task, which seems like a reasonable approach.
- However, such rules are brittle and one can readily come up with examples that will break such a rule-based system.

What is a more robust way of solving this task?

What makes the first task trivial and the second one tricky to program?

- Mathematical function for addition of two numbers, $c = a + b$, is ***already available!***
- In case of the second task of handwritten digit recognition, we ***do not have an exact mathematical function*** that takes an image as an input and outputs the recognized digit.
- With ML, we hope to uncover the mathematical form that maps images to the digits.

What makes the first task trivial and the second one tricky to program?

- Mathematical function for addition of two numbers, c = a + b, is ***already available!***
- In case of the second task of handwritten digit recognition, we ***do not have an exact mathematical function*** that takes an image as an input and outputs the recognized digit.
- With ML, we hope to uncover the mathematical form that maps images to the digits.

What makes the first task trivial and the second one tricky to program?

- Mathematical function for addition of two numbers, c = a + b, is ***already available!***
- In case of the second task of handwritten digit recognition, we ***do not have an exact mathematical function*** that takes an image as an input and outputs the recognized digit.
- With ML, we hope to uncover the mathematical form that maps images to the digits.

What makes the first task trivial and the second one tricky to program?

- Mathematical function for addition of two numbers, c = a + b, is ***already available!***
- In case of the second task of handwritten digit recognition, we ***do not have an exact mathematical function*** that takes an image as an input and outputs the recognized digit.
- With ML, we hope to uncover the mathematical form that maps images to the digits.

Tasks solvable with ML

- As a general rule, there are many tasks that humans do well, but are unable to elaborate the step by step process that they follow for solving those tasks.
- Such tasks are amenable to ML.
- We can at least attempt to address them through ML.

Tasks solvable with ML

- As a general rule, there are many tasks that humans do well, but are unable to elaborate the step by step process that they follow for solving those tasks.
- Such tasks are amenable to ML.
- We can at least attempt to address them through ML.

Tasks solvable with ML

- As a general rule, there are many tasks that humans do well, but are unable to elaborate the step by step process that they follow for solving those tasks.
- Such tasks are amenable to ML.
- We can at least attempt to address them through ML.

Tasks solvable with ML

- As a general rule, there are many tasks that humans do well, but are unable to elaborate the step by step process that they follow for solving those tasks.
- Such tasks are amenable to ML.
- We can at least attempt to address them through ML.

How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- Can we replicate this process for computers?
- Can we train computers to recognize digits in this way?
- Maybe yes!

How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- Can we replicate this process for computers?
- Can we train computers to recognize digits in this way?
- Maybe yes!

How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- Can we replicate this process for computers?
- Can we train computers to recognize digits in this way?
- Maybe yes!

How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- **Can we replicate this process for computers?**
- Can we train computers to recognize digits in this way?
- Maybe yes!

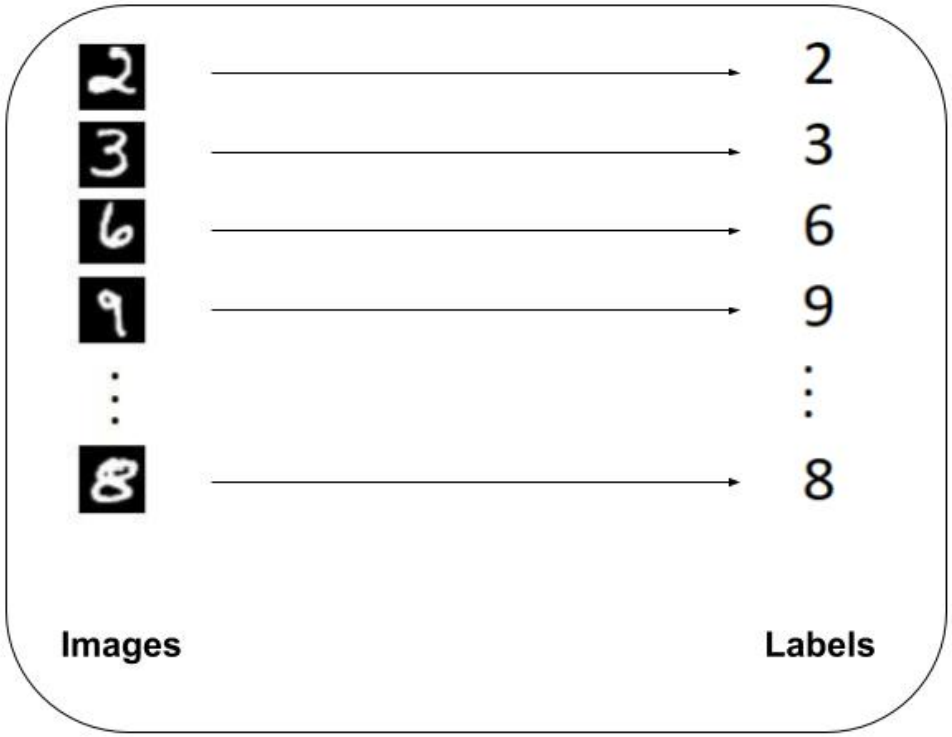
How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- **Can we replicate this process for computers?**
- **Can we train computers to recognize digits in this way?**
- Maybe yes!

How can we do that?

- Think, how do you recognize a digit from the image?
- You learnt this in school and since then, we have looked at many digits and have become better at the task.
- **Can we replicate this process for computers?**
- **Can we train computers to recognize digits in this way?**
- **Maybe yes!**

Training Data



Training Data

Model: Mathematical relationship

- Our job is to take this training data and uncover the mathematical relationship between pixels in the image and the label.
- For example, a simple linear relationship between pixels and labels is as following:

$$w_0 + w_1 \cdot pixel_1 + w_2 \cdot pixel2 + \cdots + w_m \cdot pixel_m$$

- Weights parameters: $w_0, w_1, w_2, \ldots, w_m$
- Our job is to find these parameters. How can we do it?

Model: Mathematical relationship

- Our job is to take this training data and uncover the mathematical relationship between pixels in the image and the label.
- For example, a simple linear relationship between pixels and labels is as following:

$$w_0 + w_1 \cdot pixel_1 + w_2 \cdot pixel2 + \cdots + w_m \cdot pixel_m$$

- Weights parameters: $w_0, w_1, w_2, \ldots, w_m$
- Our job is to find these parameters. How can we do it?

Model: Mathematical relationship

- Our job is to take this training data and uncover the mathematical relationship between pixels in the image and the label.
- For example, a simple linear relationship between pixels and labels is as following:

$$w_0 + w_1 \cdot pixel_1 + w_2 \cdot pixel2 + \cdots + w_m \cdot pixel_m$$

- Weights parameters: $w_0, w_1, w_2, \ldots, w_m$
- Our job is to find these parameters. How can we do it?

Model: Mathematical relationship

- Our job is to take this training data and uncover the mathematical relationship between pixels in the image and the label.
- For example, a simple linear relationship between pixels and labels is as following:

$$w_0 + w_1 \cdot pixel_1 + w_2 \cdot pixel2 + \cdots + w_m \cdot pixel_m$$

- Weights parameters: $w_0, w_1, w_2, \ldots, w_m$
- Our job is to find these parameters. How can we do it?

Model: Mathematical relationship

- Our job is to take this training data and uncover the mathematical relationship between pixels in the image and the label.
- For example, a simple linear relationship between pixels and labels is as following:

$$w_0 + w_1 \cdot pixel_1 + w_2 \cdot pixel2 + \cdots + w_m \cdot pixel_m$$

- Weights parameters: $w_0, w_1, w_2, \ldots, w_m$
- Our job is to find these parameters. How can we do it?

Finding parameters: Loss function

- Can these parameters be assigned some random numbers to get a model? Would that model be good enough? How do we measure it?
- **Loss function**: It measures disagreement between the actual outcome and the predicted outcome by applying the model to the input.
- The ideal set of parameters result in the minimum loss.
- Optimization procedure: search over the parameter space and find the optimal parameters that minimize the loss function.

Finding parameters: Loss function

- Can these parameters be assigned some random numbers to get a model? Would that model be good enough? How do we measure it?
- **Loss function**: It measures disagreement between the actual outcome and the predicted outcome by applying the model to the input.
- The ideal set of parameters result in the minimum loss.
- Optimization procedure: search over the parameter space and find the optimal parameters that minimize the loss function.

Finding parameters: Loss function

- Can these parameters be assigned some random numbers to get a model? Would that model be good enough? How do we measure it?
- **Loss function**: It measures disagreement between the actual outcome and the predicted outcome by applying the model to the input.
- The ideal set of parameters result in the minimum loss.
- Optimization procedure: search over the parameter space and find the optimal parameters that minimize the loss function.

Finding parameters: Loss function

- Can these parameters be assigned some random numbers to get a model? Would that model be good enough? How do we measure it?
- **Loss function**: It measures disagreement between the actual outcome and the predicted outcome by applying the model to the input.
- The ideal set of parameters result in the minimum loss.
- Optimization procedure: search over the parameter space and find the optimal parameters that minimize the loss function.

Finding parameters: Loss function

- Can these parameters be assigned some random numbers to get a model? Would that model be good enough? How do we measure it?
- **Loss function**: It measures disagreement between the actual outcome and the predicted outcome by applying the model to the input.
- The ideal set of parameters result in the minimum loss.
- Optimization procedure: search over the parameter space and find the optimal parameters that minimize the loss function.

Evaluation

- Once the model is trained, it is evaluated with appropriate metrics like accuracy.
- If this metric is satisfactory, we have a model that can be used for making predictions.
- In case, it is not satisfactory, the process is repeated by adding more features or more data, depending on the metric.

Evaluation

- Once the model is trained, it is evaluated with appropriate metrics like accuracy.
- If this metric is satisfactory, we have a model that can be used for making predictions.
- In case, it is not satisfactory, the process is repeated by adding more features or more data, depending on the metric.

Evaluation

- Once the model is trained, it is evaluated with appropriate metrics like accuracy.
- If this metric is satisfactory, we have a model that can be used for making predictions.
- In case, it is not satisfactory, the process is repeated by adding more features or more data, depending on the metric.

Evaluation

- Once the model is trained, it is evaluated with appropriate metrics like accuracy.
- If this metric is satisfactory, we have a model that can be used for making predictions.
- In case, it is not satisfactory, the process is repeated by adding more features or more data, depending on the metric.

Fundamental Assumption

Why does this method of learning parameters from training data work? What are some of the fundamental assumptions that we make in ML?

The training and test data would come from the same distribution.

Fundamental Assumption

Why does this method of learning parameters from training data work? What are some of the fundamental assumptions that we make in ML?

The training and test data would come from the same distribution.

Fundamental Assumption:Example of History test

- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

Fundamental Assumption:Example of History test

- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

Fundamental Assumption:Example of History test

- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

Fundamental Assumption:Example of History test

- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

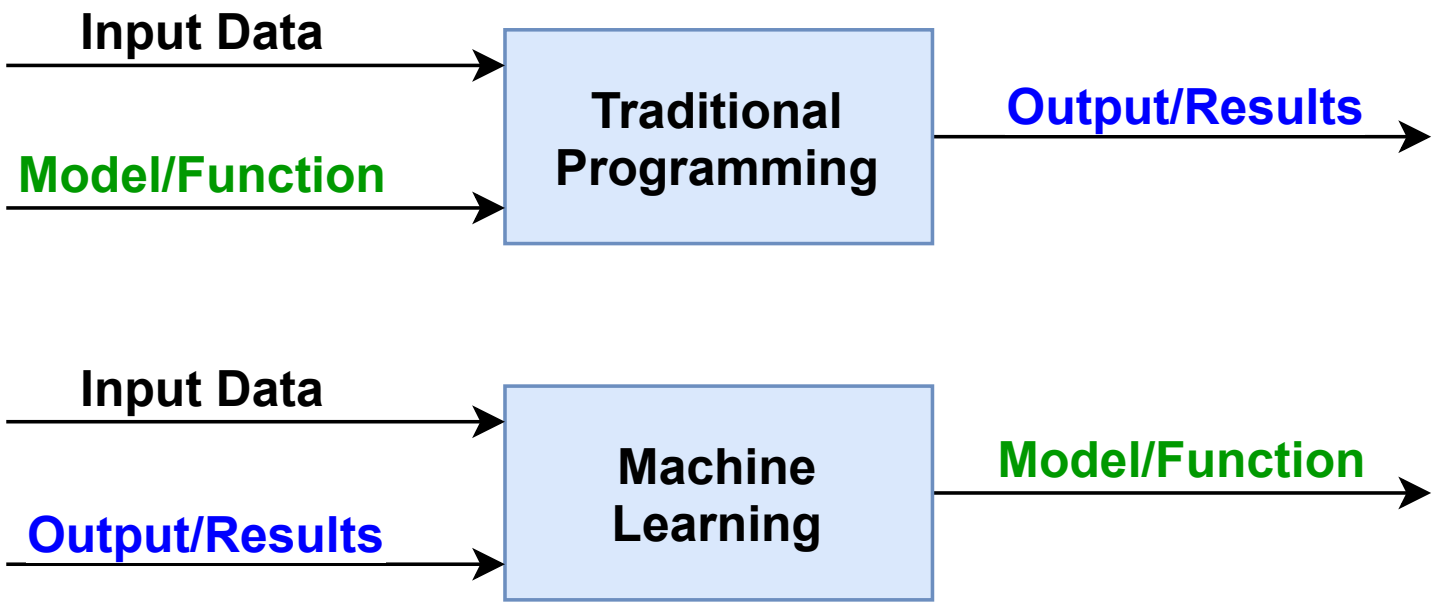
Fundamental Assumption:Example of History test

- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

Fundamental Assumption:Example of History test

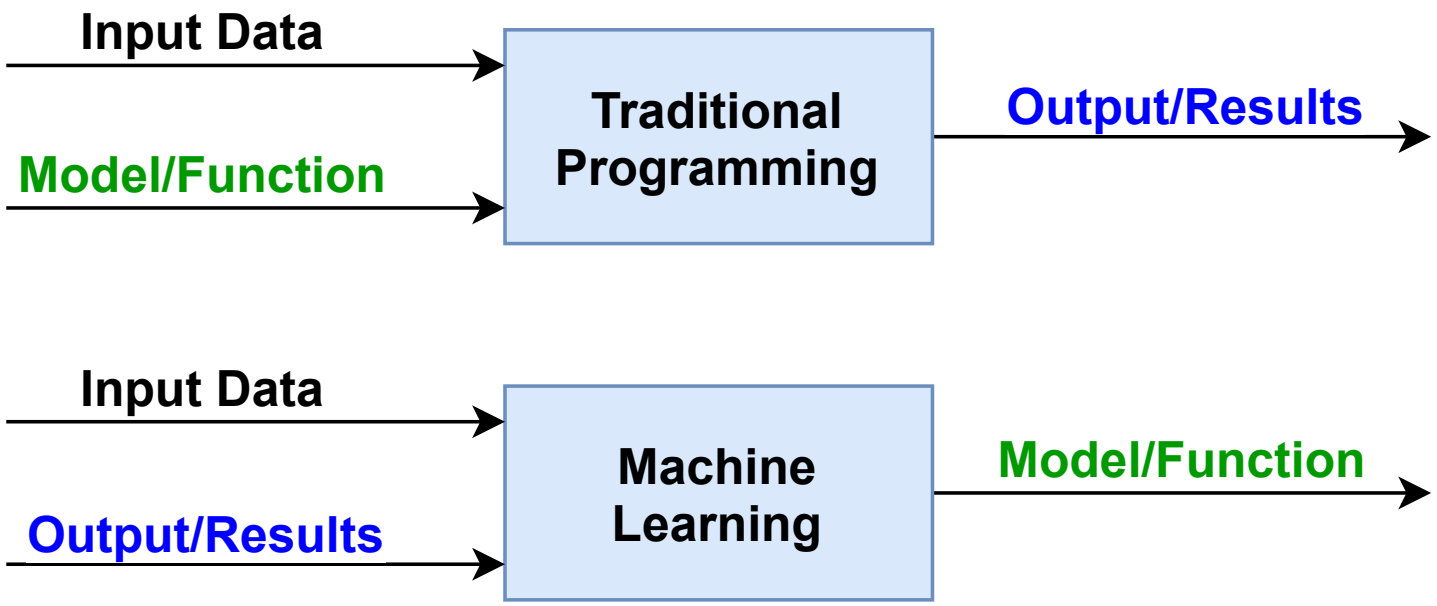
- Students study history syllabus.
- Training data and test data are from the same distribution, if the teacher asks questions from this syllabus.
- If the questions are not from the syllabus - then training and test data do not come from the same distribution.
- Students perform better if the exam paper is based on the syllabus that they study or train on.
- Students do not perform well if questions are NOT from the syllabus.

Summary



- Once the model or rule is learnt, traditional programming can be used to get output for a given input label.

Summary



- Once the model or rule is learnt, traditional programming can be used to get output for a given input label.