Machine Learning – Day 1 Notes

Types of Learning in Machine Learning

---

1. Initial Understanding

Initially, I believed that **Machine Learning was mainly about selecting the right algorithm**.
This view changed after understanding how **different systems learn from data**.

2. Core Insight

Machine Learning is **not a single concept**.

It is defined by **how a system learns from data**, not just by the algorithm it uses.
The **learning approach** determines:

- How data is prepared
- What kind of feedback is available
- Which algorithms are suitable

The type of learning matters more than the algorithm itself.

3. Types of Machine Learning

A. Supervised Learning

Learning with labeled data

In supervised learning:

- The dataset contains **inputs with correct answers (labels)**
- The model learns by comparing its predictions with the actual answers
- Errors are used to improve performance

Real-life examples:

- **Email spam detection**
  Emails are labeled as “Spam” or “Not Spam”
- **House price prediction**
  Past house data includes size, location, and actual selling price
- **Medical diagnosis systems**
  Patient data with known disease outcomes

Key idea:
The model learns with guidance, similar to a student learning from a teacher.

B. Unsupervised Learning

Learning without labeled data

In unsupervised learning:

- No correct answers are provided
- The system identifies **patterns, similarities, or structures**
- The goal is understanding data rather than prediction

Real-life examples:

- **Customer segmentation in marketing**
  Grouping customers based on purchase behavior
- **Organizing photos on a smartphone**
  Automatically grouping similar faces or scenes
- **Market basket analysis**
  Finding products that are frequently bought together

Key idea:
The model explores data on its own and discovers hidden patterns.

C. Reinforcement Learning

Learning through interaction and feedback

In reinforcement learning:

- The model takes actions in an environment
- Each action results in a **reward or penalty**
- The goal is to maximize long-term rewards

Real-life examples:

- **Game-playing AI** (chess, video games)
  Learning strategies by winning or losing
- **Self-driving cars**
  Learning safe driving behavior through continuous feedback
- **Robotics**
  Learning how to walk or pick objects through repeated attempts

Key idea:
The system improves by learning from experience and consequences.

4. Shift in Perspective

Earlier View

Machine Learning is mostly about:

- Choosing algorithms like Linear Regression, KNN, or Decision Trees

Updated Understanding

Machine Learning starts with:

- Understanding the **problem**
- Identifying the **type of learning**
- Then selecting an appropriate algorithm

The algorithm is a **tool**, not the foundation.

5. Final Takeaway

Understanding **how learning happens** makes Machine Learning:

- More logical
- More structured
- Less intimidating

Learning approach → Data → Algorithm → Model

Machine Learning – Day 3 Notes

Why Machine Learning Systems Fail

---

1. Initial Assumption

I assumed that **Machine Learning fails because models are weak or incorrect**.

This understanding turned out to be misleading.

2. Key Realization

Most Machine Learning problems **do not fail at the model stage**.
They fail **much earlier in the pipeline**.

Even a highly advanced model cannot succeed if earlier steps are flawed.

3. Machine Learning Is a Process

Machine Learning is **not a shortcut** where data is fed into a model to get results.

It is a **step-by-step process**, where:

- Each stage depends on the correctness of the previous one
- Skipping or rushing any step silently damages the final outcome

4. Common Reasons ML Projects Fail

A. Unclear Problem Definition

If the problem is not clearly defined:

- The model may optimize the wrong objective
- Success metrics become meaningless

Real-life example:

- Trying to "predict customer behavior" without specifying whether the goal is:
    - Increasing retention
    - Improving recommendations
    - Reducing churn

Insight:
A model cannot solve a problem that humans have not clearly framed.

B. Poor Data Quality

Data issues are one of the **most common failure points** in ML.

Problems include:

- Missing values
- Incorrect labels
- Noisy or biased data
- Inconsistent formats

Real-life example:

- Predicting house prices using outdated or incorrect property data
- Training a spam detector with wrongly labeled emails

Insight:
Better data often improves results more than a better algorithm.

C. Wrong Evaluation Strategy

If evaluation is incorrect:

- The model may appear accurate but fail in real-world use
- Decisions based on results become unreliable

Real-life example:

- Using accuracy for an imbalanced dataset where most values belong to one class
- Testing on data too similar to training data

Insight:
A model is only as good as how it is evaluated.

5. Why Better Models Don't Fix These Issues

- A powerful model cannot compensate for:
  - Poor problem framing
  - Messy or misleading data
  - Incorrect evaluation metrics

This explains why many ML projects fail **even with advanced algorithms**.

6. Shift in Understanding

Earlier Thinking

Machine Learning is mainly about:

- Training models
- Improving algorithms
- Increasing accuracy

Updated Thinking

Machine Learning is about:

- Clear thinking from start to end
- Treating ML as a **system**, not just a model
- Ensuring every step is logically sound
-

7. Final Takeaway

Machine Learning success depends on the entire pipeline:

Problem Definition → Data Quality → Feature Design → Model → Evaluation → Deployment

Skipping any step quietly breaks everything that follows.

Machine Learning is less about "training models"
and more about **thinking clearly throughout the process**.

Machine Learning – Day 4 Notes

Why Data Quality Matters More Than Data Quantity

---

1. Initial Assumption

I believed that **better Machine Learning results require more data**.

This assumption turned out to be incomplete and often misleading.

2. Key Realization

More data **does not automatically improve model performance**.

What matters far more is:

- How accurate the data is
- How relevant it is to the problem
- Whether it represents reality correctly

Data quality has a greater impact than data quantity.

3. Understanding Data Quality in Machine Learning

A. Noisy Data

Noisy data contains:

- Errors
- Random variations
- Inconsistent or incorrect values

Impact:

- True patterns become harder to detect
- Models learn randomness instead of signal

Real-life example:

- Sensor data with faulty readings
- User-entered data with spelling mistakes or incorrect values

Insight:
Noise hides meaningful patterns, regardless of dataset size.

B. Biased Data

Biased data represents the world **unevenly or unfairly**.

Impact:

- Models learn incorrect relationships
- Predictions become unreliable or unfair

Real-life example:

- Hiring models trained on historically biased recruitment data
- Credit risk models trained on limited demographic groups

Insight:
Models only learn what data shows — not what is true.

C. Irrelevant Data

Irrelevant data does not contribute to solving the target problem.

Impact:

- Adds complexity without improving understanding
- Can increase confidence while reducing real accuracy

Real-life example:

- Including a user's device color when predicting loan approval
- Adding unrelated website metrics to sales prediction

Insight:
More features do not mean more information.

4. Small Clean Data vs Large Messy Data

A **small, well-curated dataset** can outperform a large dataset when:

- Labels are accurate
- Features are meaningful
- Noise and bias are controlled

Why this happens:

- Clean data provides clearer learning signals
- Models generalize better with trustworthy inputs

5. Shift in Perspective

Earlier Thinking

- More data → better model
- Focus on collecting as much data as possible

Updated Thinking

- Better data → better learning
- Focus on:
  - Data cleaning
  - Data validation
  - Data relevance

6. Practical ML Mindset Change

Before asking:

“Which model should I use?”

It is more important to ask:

“Can this data be trusted?”

Data understanding comes **before** model selection.

7. Final Takeaway

Machine Learning performance depends heavily on data quality.

Clean data → Clear patterns → Reliable models

Large datasets cannot compensate for:

- Noise
- Bias
- Irrelevance

In Machine Learning,
quality creates accuracy — not quantity.

Machine Learning – Day 5 Notes

The Importance of Features in Machine Learning

---

1. Initial Thinking

I focused heavily on **choosing better algorithms**, assuming stronger models would automatically lead to better performance.

This perspective changed after understanding the role of **features**.

2. Key Realization

An algorithm **does not understand reality**.
It only reacts to **what is given to it as input**.

If the features fail to represent the real problem:

- Model performance drops
- Even advanced algorithms cannot recover

The model learns only from the information it receives.

3. What Are Features?

Features are the **inputs** provided to a Machine Learning model.

They represent:

- Observations about the real world
- Information the model uses to find patterns and make predictions

Features define what the model is capable of learning.

4. Weak Features and Their Impact

A. Weak or Poorly Designed Features

Weak features:

- Do not reflect the true drivers of the problem
- Contain noise or vague signals
- Are loosely related to the target outcome

Impact:

- Important patterns remain hidden
- The model struggles to learn meaningful relationships

Real-life example:

- Predicting house prices using only the number of windows
- Predicting student performance using seat position in class

Insight:
No algorithm can extract useful insight from irrelevant or weak information.

5. Strong Features and Their Advantage

B. Well-Designed, Meaningful Features

Strong features:

- Capture real-world relationships
- Reduce ambiguity in the data
- Make patterns easier to learn

Impact:

- The problem becomes simpler before training starts
- Even simple models perform well

Real-life example:

- House price prediction using location, area, and number of rooms
- Credit risk prediction using income stability and repayment history

Insight:
Good features reduce the burden on the algorithm.

6. Why Feature Design Comes Before Model Choice

- Algorithms optimize patterns **within features**
- They cannot invent missing information
- Feature quality sets the upper limit on model performance

This explains why:

- Simple models with strong features can outperform
- Complex models with weak features

7. Shift in Perspective

Earlier Focus

- "Which algorithm should I use?"
- "How complex should my model be?"

Updated Focus

- "What information am I giving the model?"
- "Do these features truly represent reality?"

8. Final Takeaway

Machine Learning performance is heavily constrained by feature quality.

Reality → Features → Algorithm → Predictions

If reality is poorly captured at the feature level,
no algorithm can compensate for it.

Strong features simplify the problem
before learning even begins.

Machine Learning – Day 6 Notes

Understanding Underfitting and Overfitting

---

1. Initial Assumption

I believed that **mistakes in Machine Learning mainly come from weak or simple models**.

I learned that this is only part of the picture.

2. Key Realization

A model can fail in **two fundamentally different ways**:

1. By **not learning enough**
2. By **learning too much**

Both lead to poor performance, but for opposite reasons.

3. Two Types of Model Learning Failures

A. Underfitting

When the model learns too little

Underfitting happens when:

- The model is too simple
- Important patterns in the data are ignored
- The model fails to capture relationships that actually matter

Impact:

- Poor performance on both training and new data
- Predictions are overly general and inaccurate

Real-life example:

- Predicting house prices using only the number of rooms
- Predicting exam scores using only attendance

Core insight:
The model lacks the capacity or information to learn meaningful patterns.

B. Overfitting

When the model learns too much

Overfitting happens when:

- The model memorizes training data
- Noise is mistaken for real patterns
- Performance is excellent on training data but poor on new data

Impact:

- High training accuracy
- Low real-world reliability

Real-life example:

- Memorizing answers instead of understanding concepts
- A model that fits historical stock data perfectly but fails on future prices

Core insight:
The model learns patterns that do not generalize to real situations.

4. Why This Changes How Performance Issues Are Viewed

When a model performs poorly, the issue may not be:

- The algorithm choice
- The lack of model complexity

Instead, the problem may be:

- Too simple learning (underfitting)
- Too specific learning (overfitting)

Understanding *how* the model is learning becomes more important than *which* model is used.

5. Balance Is the Goal

Effective Machine Learning aims to:

- Capture meaningful patterns
- Ignore noise and randomness
- Generalize well to unseen data

This balance lies **between underfitting and overfitting**.

6. Shift in Perspective

Earlier Thinking

- Poor results mean the model is weak

Updated Thinking

- Poor results may indicate:
  - Insufficient learning
  - Excessive memorization

Diagnosis matters more than assumptions.

7. Final Takeaway

Machine Learning performance issues are often **learning issues**, not algorithm issues.

Data → Learning Behavior → Performance

Understanding whether a model is:

- Ignoring patterns
- Or memorizing noise

is key to improving results.

Good models don't learn everything —
they learn what matters.