ML Concepts - Generalization

What is Generalization?

- Generalization refers to **your model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model**.

Figure 1. Sick (blue) and healthy (orange) trees.

- Figure 1 depicts around 50 dots, evenly split between blue (sick) and orange (healthy) trees.
- The southwest quadrant is primarily populated with orange dots, though a few venture briefly into the other quadrants. Blue dots dominate the northeast quadrant, with some spilling into other areas.

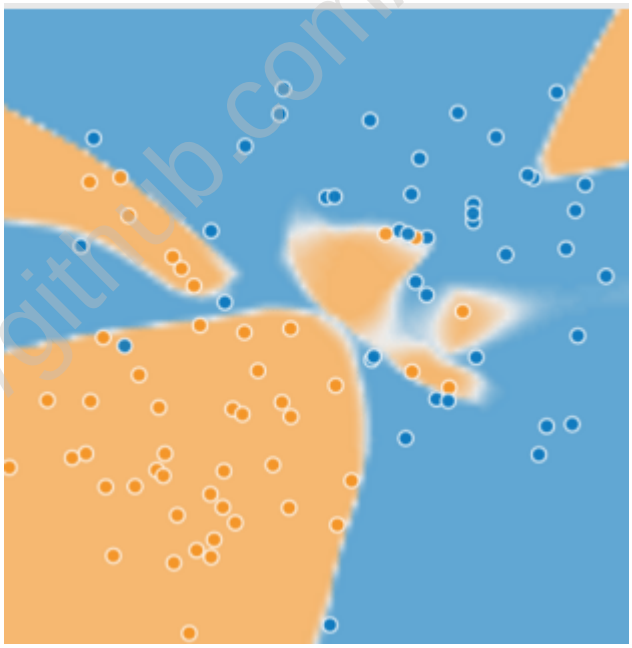

Figure 2. A complex model for distinguishing sick from healthy trees.

- This figure introduces a machine learning model that effectively separates sick trees from healthy ones, achieving a very low loss.
- The model accurately encloses most of the blue and orange dots with a collection of complex shapes.

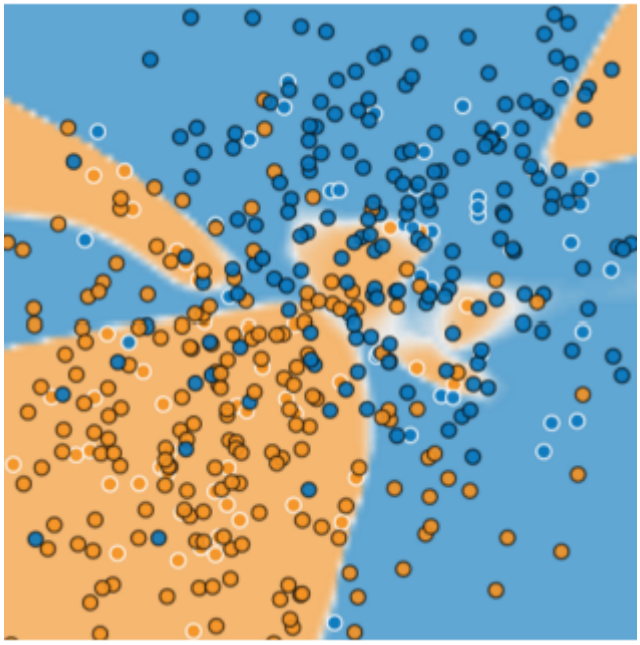

Figure 3. The model did a bad job predicting new data.

- Now, with an additional 100 dots, demonstrates that the model poorly adapts to new data, mis categorizing many points.
- Many new dots fall well outside of the predicted model, indicating potential overfitting.
- The model in Figures 2 and 3 overfits the training data, resulting in low training loss but poor performance on new data.
- Overfitting occurs when a model tries to fit the training data too closely, potentially leading to poor generalization on new data.
- **Balancing Complexity:** Overfitting is primarily caused by making a model overly complex for the given dataset.
- William of Ockham's principle of simplicity is applied in machine learning, favoring simpler models over overly complex ones.
- In ML terms, this means that a less complex model is more likely to produce reliable results not solely based on the peculiarities of the sample.
- Three basic assumptions guide generalization:
  - Examples are drawn independently and identically (i.i.d) at random from the distribution, meaning examples don't influence each other.
  - The distribution is stationary, indicating it doesn't change within the dataset.
  - Examples are drawn from partitions of the same distribution.
- Violating these assumptions can impact model performance.