Supervised Learning

Regression

Classification

Ensemble Techniques

Neural Networks

Julia Lenc

Content

Introduction

- What is Supervised Learning?
- Business appications
- Classification and Regression

Core methods

- Classification (logistic regression, Naive Bayes, k-NNs)
- Regression (linear, non-linears, time series)

Advanced methods

- Decision Tree
- Support Vector Machines

Ensemble methods

- Bagging, Boosting and Stacking

Neural Networks

- FFNN, CNN, RNN, Transformers

Methods summary: from simplest to most complex

Julia Lenc

What is Supervised Learning?

Definition

Supervised learning is a group of methods that learn from historical data with known outcomes to predict future values or classify new observations. It enables targeted decisions, resources and performance optimization.

Areas

🧩 Value Creation: predict churn or satisfaction and explain their drivers.

💰 Revenue Model: sales forecast, customer lifetime value (CLV), long-term pricing strategies, dynamic pricing.

🔍 Market Opportunity: market size forecast, pockets of growth, stores or leads potential, market research simulation (synthetic consumer).

🛒 Go-to-Market: price and promotional response modeling, evaluate channel performance, product recommenders.

🏢 Operations: demand forecast, risks and delays flagging.

Julia Lenc

Core Methods

Always consider them first!

Julia Lenc

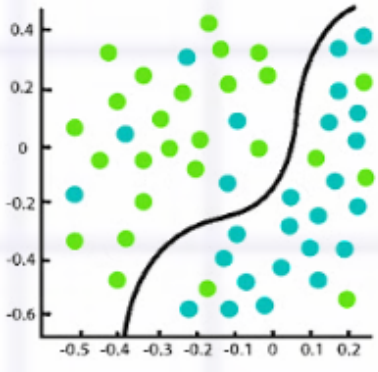
Supervised Learning Types

Classification

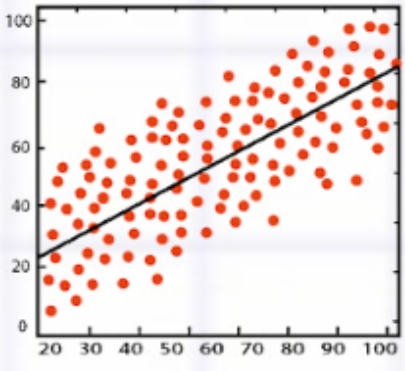
Assignment of a probability or a category to an observation.
Key questions: "Which group?", "How likely?", “Is it A or B?”
Popular applications: churn prediction, recommender systems

Regression

Prediction of a continuous value - number or quantity.
Key questions: "How much?", "What will the value be?"
Popular applications: sales forecast, budget changes, pricing



Classification



Regression

Julia Lenc

Core Classification Methods

Logistic Regression

The likelihood of belonging to a category. Highly interpretable.
Coefficients show the impact of variables on the prediction.
Popular applications: churn, recommenders.

Naive Bayes

The most likely class. Predicts be evaluating the relationship between the features of an observation and class.
Popular applications: text analysis (sentiment, spam/not spam).

k-NNs (k-nearest neighbours)

The most likely class. No need for a training or assumptions assumptions about data distribution.
Popular applications: targeted marketing, recommenders.

Julia Lenc

Data Requirements

	Logistic Regression	Naive Bayes	k-NNs
Independence of errors	🔴	🔴	🟡
Independence of features	🔴	🟡	🟡
No outliers	🟡	🟡	🔴

Legend:

🔴 Must-Have: essential for the method to give valid results.

🟡 Desirable: desirable for better performance or interpretation.

Julia Lenc

Core Regression Methods

Logistic Regression

Future continuous value. Linear relationship between variables. Highly interpretable. Coefficients show the impact of variables on the prediction.
Popular applications: sales forecasting, budget changes.

Non-Linear Regression

Future continuous value. Non-linear but defined relationships (polynomial, quantile, Poisson). Interpretable in simple cases.
Popular applications: growth modeling, forecast by segment.

Time Series

Future continuous value. Uses trends, seasonality, past values.
For sequential data and time-dependent domains.
Popular applications: stock price, demand, website traffic.

Julia Lenc

Data Requirements

	Linear Regression	Non-Linear Regression	Time Series
Linearity	🔴	🟡	🟡
Normal distribution	🟡	🟡	🟡
No outliers	🟡	🟡	🟡
Independence of errors	🟡	🟡	🟡
Constant spread of errors	🔴	🔴	🟡
Independence of features	🔴	🔴	na
Stationarity	na	na	🔴

Legend:

🔴 Must-Have: essential for the method to give valid results.

🟡 Desirable: desirable for better performance or interpretation.

Julia Lenc

Advanced Methods

Data doesn’t meet requirements

Usuall more robust and flexible predictions

Julia Lenc

Advanced Methods

Decision Tree

Splits data into branches (decisions) based on feature thresholds to reach predictive outcomes. For classification and regression tasks. Interpretatble. Prone to overfitting.
Popular applications: pricing analysis with cliffs, key impact factors, customer journey, most likely to churn customers.

Support Vector Machines

Finds the best boundary - hyperplane - that separates data points into distinct classes (SVM) or predicts continuous values (SVR) by using support vectors. Rather difficult to interpret. Prone to overfitting.
Popular applications: pricing analysis with cliffs, churn, fraud detection, insurance premiums calculations.

Julia Lenc

Ensembles

Enhanced predictive performace

at expence of

simplicity and interpretability

Julia Lenc

Emsemble Methods

Bagging (Random Forest)

Reduces variance by combining predictions from multiple models trained on different subsets of data (bootstrapping). Moderate interpretability. Less prone to overfitting.

Popular applications: portfolio risk, churn, fraud detection.

Boosting (Ada Boost, Gradient Boosting Machines, XG Boost)

Reduces bias by sequentially training models, where each focuses on correcting errors made by previous ones. High performance but low interpretability. Prone to overfitting.

Popular applications: credit risk score, campaing optimization

Bagging + Boosting = **Stacking**

Julia Lenc

Neural Networks

The data is complex that even Ensembles fall short

The data is unstructured

Julia Lenc

Neural Networks

Feed Forward Neural Network (FFNN)

Processes input data through layers in one direction. No loops. For structured numerical data and tabular datasets.

Popular applications: predictive modeling (demand, churn, fraud), creding risk scoring, investment portfolio optimization.

Convolutional Neural Network (CNN)

For data with spatial hierarchies, such as images. Can be adapted for structured/numerical data using grids or filters.

Popular applications: shelf image analysis for stock monitoring, telecom network monitoring for anomaly detection.

Julia Lenc

Neural Networks

Recurrent Neural Network (RNN)

For sequential data. Captures time-based dependencies. Specialized versions include LSTM and GRU (long-term dependencies). Often replaced with Transformers.

Popular applications: time series forecast, customer behavior prediction in telecom or banking.

Transformers

For tasks requiring rich context modeling: language, sequential or interaction data. Handles hem thanks to its attention mechanism and parallel processing.

Popular applications: recommender systems, marketing targeting, chatbots and virtual assistants.

Julia Lenc

Escalation Path

Classification:
Logistic Regression
Naive Bayes
k-NNs

Regression:
Time Series
Linear Regression
Non-Linear Regression

Advanced Methods:
Decision Tree
Support Vector Machines

Emsemble Methods:
Bagging and Boosing

Neural Networks

Julia Lenc

Coming next

1. The foundation

- Trade-offs in predictive modeling
- Data preparation (cleansing and encoding)

2. Core and Advanced methods

- Classifications
- Regressions
- Decision Tree and Support Vector Machines

3. Ensemble techniques

- Bagging
- Boosting
- Stacking

4. Neural Networks

- Feed Forward Neural Network (FFNN)
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN) and Transformers

Julia Lenc

Did you find it useful?

Save

Share

Follow

Analytics, Market Research,
Machine Learning and AI

Julia Lenc