Makine Öğrenmesi
(Machine Learning)

AU, 2018
Bölüm 1: 
Makine Öğrenmesi
Machine Learning

• **Machine learning** is the subfield of computer science concerned with computational learning. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

https://en.wikipedia.org/wiki/Machine_learning

• **Machine learning**, in artificial intelligence (a subject within computer science), discipline concerned with the implementation of computer software that can learn autonomously.

https://global.britannica.com/technology/technology/machine-learning
Machine Learning

Kendi başına öğrenebilen yazılımların gerçekleştirimi ile ilgili disiplin.
Machine Learning

• "A computer program is said to learn from experience $E$
• with respect to some class of tasks $T$
• and performance measure $P$ if its performance at tasks in $T$, as measured by $P$,
• improves with experience $E$. " Tom Mitchell (wiki)
Books

• Python Machine Learning 1st Edition
  by Sebastian Raschka

• Data Science from Scratch: First Principles with Python
  Joel Grus, "O'Reilly Media, Inc.", 2015, 330 sayfa

• http://scikit-learn.org/0.16/_downloads/scikit-learn-docs.pdf
• **Supervised learning**: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.

• **Unsupervised learning**: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

• **Reinforcement learning**: A computer program interacts with a dynamic environment in which it must perform a certain goal (araba sürme, rakibe karşı oyun oynama, ...). The program is provided feedback in terms of rewards and punishments as it navigates its problem space. [http://www.wikiwand.com/en/Machine_learning](http://www.wikiwand.com/en/Machine_learning)
1. Supervised Learning - I
(Öğreticili Öğrenme, Danışmanlı Öğrenme, Gözetimli Öğrenme, ...)

• I - Regresyon (Regression) https://blog.udemy.com/linear-regression-example/

1. Supervised Learning – II

I - Regresyon (Regression) [https://blog.udemy.com/linear-regression-example/](https://blog.udemy.com/linear-regression-example/)

The Equation for regression can be expressed as follows:

Regression Equation(y) = a + bx

Slope(b) = \((N\Sigma XY – (\Sigma X)(\Sigma Y)) / (N\Sigma X^2 – (\Sigma X)^2)\)

Intercept(a) = \((\Sigma Y – b(\Sigma X)) / N\)

Where:

- \(x\) and \(y\) are the variables.
- \(b\) = The slope of the regression line
- \(a\) = The intercept point of the regression line and the y axis.
- \(N\) = Number of values or elements
- \(X\) = First Score
- \(Y\) = Second Score
- \(\Sigma XY\) = Sum of the product of first and Second Scores
- \(\Sigma X\) = Sum of First Scores
- \(\Sigma Y\) = Sum of Second Scores
- \(\Sigma X^2\) = Sum of square First Scores

Simple linear regression is a technique that displays the relationship between variable “\(y\)” based on the values of variable “\(x\)”.

Reklam Harcamaları ve Satışlar arası ilişki:
Regresyon (Bağlanım) Tanımı ve Çeşitleri - I

**Regresyon analizi**, iki ya da daha çok değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur.

Regresyonda, değişkenlerden biri bağımlı diğerleri bağımsız değişken olmalıdır.

Eğer tek bir değişken kullanılarak analiz yapıyorsa buna **tek değişkenli regresyon**, birden **çoğunok değişken kullanılıyor**sa çok değişkenli regresyon analizi olarak isimlendirilir (The case of one explanatory variable is called *simple linear regression*. For more than one explanatory variable, the process is called *multiple linear regression*). (wikipedia)
Doğrusal Regresyon (Linear Regression): In statistics, linear regression is a linear approach for modelling the relationship between a scalar dependent variable $y$ and one or more explanatory variables (or independent variables) denoted $X$.

Doğrusal Olmayan Regresyon: In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. (wikiwand)
1. Supervised Learning - III

• II – Classification (Sınıflandırma) Örnek 1
Sıcaklık, nem, rüzgar ve hava durumuna göre tenis oynayıp oynamamaya karar verme. İki adet sınıf ton (yes, no) birisine karar verilmeli. 4 adet öznitelik ve 1 adet çıktı var.

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Windy</th>
<th>Play</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>hot</td>
<td>high</td>
<td>false</td>
<td>no</td>
</tr>
<tr>
<td>sunny</td>
<td>hot</td>
<td>high</td>
<td>true</td>
<td>no</td>
</tr>
<tr>
<td>overcast</td>
<td>hot</td>
<td>high</td>
<td>false</td>
<td>yes</td>
</tr>
<tr>
<td>rainy</td>
<td>cool</td>
<td>normal</td>
<td>false</td>
<td>yes</td>
</tr>
<tr>
<td>rainy</td>
<td>cool</td>
<td>normal</td>
<td>true</td>
<td>no</td>
</tr>
<tr>
<td>overcast</td>
<td>cool</td>
<td>normal</td>
<td>true</td>
<td>yes</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>high</td>
<td>true</td>
<td>no</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>high</td>
<td>false</td>
<td>no</td>
</tr>
</tbody>
</table>

Örnek 2:
Otomobilleri 4’e sınıflandırma

Attribute Values: (6 öznitelik)

- buying       v-high, high, med, low
- maint        v-high, high, med, low
- doors        2, 3, 4, 5-more
- persons      2, 4, more
- lug_boot     small, med, big
- safety       low, med, high

Sınıflar:
Çok İyi, İyi, Orta, Kötü


https://archive.ics.uci.edu/ml/datasets/car+evaluation
Lojistik Regresyon

• In statistics, logistic regression, or logit regression, or logit model is a regression model where the dependent variable (DV) is categorical.

1. Supervised Learning - III

• Öğreticili Öğrenme Yöntemleri
  • Karar Ağaçları (Decision Tree) Learning
  • Naive-Bayes Classifier
  • Destek Vektör Makineleri (Support Vector Machines - SVM)
  • K-En Yakın Komşu (K-Nearest Neighbor - KNN)
  • Doğrusal Regresyon (Linear Regression)
  • Lojistik Regresyon (Logistic Regression)
  • Yapay Sinir Ağları (Neural Networks) (Example: Multilayer Perceptron)
  • Rastgele Orman (Random Forest) [Ensemble Learning Kapsamında]
  • …
2. Unsupervised Learning - I
(Öğreticisiz Öğrenme, Danışmansız Öğrenme, Gözetimsiz Öğrenme, …)

- **Unsupervised machine learning** is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data (a classification or categorization is not included in the observations). (wiki)
  - **Clustering**: To group samples such that objects within the same cluster are more similar to each other than to the objects from another cluster.
  - **Association**: To discover the probability of the co-occurrence of items in a collection.
  - **Dimensionality Reduction**: True to its name, Dimensionality Reduction means reducing the number of variables of a dataset while ensuring that important information is still conveyed. (PCA gibi) (https://www.kdnuggets.com/2017/10/top-10-machine-learning-algorithms-beginners.html)
2. Unsupervised Learning - II

Clustering:

• K-Means Örneği

• Örnek: Özellikleri yönü ile birbirine benzer müşteriler.
2. Unsupervised Learning - III
Association Analysis (Birlik telik Analizi)

- Birlik telik Analizi Yöntemleri: Apriori, FP-Growth
2. Unsupervised Learning - IV

- Kümeleme Yöntemleri
  - Yoğunluğa Dayalı (Density-based): DBSCAN, OPTICS, …
  - Model Tabanlı (Model-based): Self-Organizing Map (SOM) NN, …
  - Hiyerarşik (Hierarchical Clustering) …
  - İzgara Tabanlı (Grid-based) …
  - Esnek Hesaplama Teknikleri (Soft Computing Tech.): FCM, GA
  - …
3. Reinforcement Learning - I
(Destekleyici Öğrenme, Pekiştirmeli Öğrenme)

• Satranç: Ödül oyun sonunda gelir, kazanmak.
• Masa Tenisi: Ödül, her puan alışta gelir.
• Sistem / Ortam, hareketin doğru mu yanlış mı olduğunu belirtir.
• Ödül ve Ceza mekanizması
3. Reinforcement Learning - II

- Destekleyici Öğrenme Yöntemleri
  - Temporal Difference (TD) Learning
  - Adaptive Actor-Critics
  - Q-Learning
  - Neural Networks
    - Learning Vector Quantization (LVQ)
  - Error-Driven Learning
Top 10 ML Algorithms

Data Science  
Veri Bilimi

• **Data science**, also known as **data-driven science**, is an interdisciplinary field about scientific methods, processes and systems to extract **knowledge** or insights from **data** in various forms, either structured or unstructured,[1][2] similar to **Knowledge Discovery in Databases** (KDD). (wiki)

• It employs techniques and theories drawn from many fields within the broad areas of **mathematics**, **statistics**, **information science**, and **computer science**, in particular from the subdomains of **machine learning**, **classification**, **cluster analysis**, **data mining**, **databases**, and **visualization**. (wiki)
Anomaly Detection (wiki)
Anomali Tespiti

• In data mining, anomaly detection (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset.

• Anomaly detection is applicable in a variety of domains, such as intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, ...

• Supervised vs. unsupervised anomaly detection.
Makine Öğrenmesinde Kullanılan Ortamlar

• Python
• R
• Matlab
• WEKA
• ...

Ödev: Anaconda platformunu kurun, Spyder IDE’sini simgesinden çalıştırın (veya Jupyter/Notebook). ML örnekleri bulup çalıştırın.