



دانشگاه تربیت مدرس

دانشکده مهندسی برق و کامپیوتر

دانشکده علوم و فناوری های بین رشته ای

تمرین چهارم درس یادگیری ماشین

مدل پنهان مارکوف در یادگیری ماشینی

استاد محترم درس:

جناب آقای دکتر قادری

دستیاران آموزشی:

یگانه چتری

نیلوفر مقدس

- ✓ پروژه فقط با زبان برنامه نویسی پایتون قابل قبول میباشد.
- ✓ فایل تحولی شما، یک فایل زیپ شده نهایی شامل گزارش کار (فایل pdf) و فایل کد پایتون با پسوند ipynb (Jupyter Notebook) میباشد. لطفا آن را به صورت زیر نامگذاری و ارسال نمایید.
- ✓ HW_٤_HMM_[LastName]_[FirstName]
- ✓ گزارش کار خود را در یک فایل pdf تحويل دهید و از گذاشتن صرفا اسکرین شات های پشت سرهم از کد در گزارش کار خودداری کنید.
- ✓ توجه داشته باشید که در فایل ارسالی پایتون، خروجی هر سلول (شامل نمودار، خروجی عددی و غیره) حتما ذخیره شده و قابل مشاهده باشد.
- ✓ لازم است حتما نتایج بدست آمده را گزارش و تحلیل کنید.
- ✓ علاوه بر مهارت حل سوالات، نوشتن پاسخ مینی پروژه ها در فرمت گزارش فنی (فصل بندی و صفحه بندی مناسب، رعایت اصول نگارش و ...) برای دانشجویان تحصیلات تکمیلی اهمیت دارد، این مورد نیز در ارزشیابی لحاظ می شود.
- ✓ در صورت فراموشی در ارسال کد پایتون، هیچ نمره ای به شما تعلق نخواهد گرفت.
- ✓ در صورت مشاهده تشابه در هر بخش از انجام پروژه، نمره هر دو نفر صفر لحاظ میگردد.

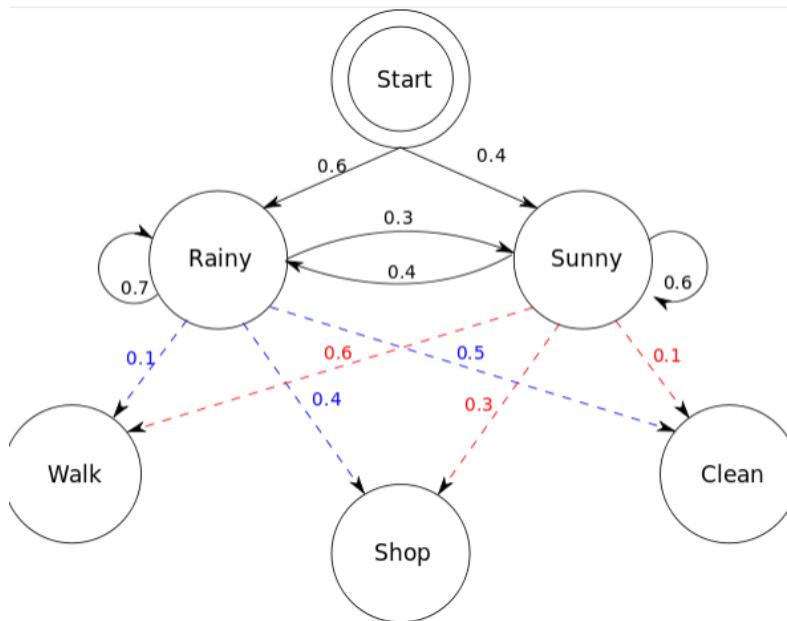
برای پاسخ به سوالات دانشجویان در مورد مینی پروژه ها، دو راه ارتباطی وجود دارد:

1. ایمیل: برای پرسش سوال از طریق ایمیل، در قسمت To، ایمیل دستیار آموزشی این مینی پروژه، خانم یگانه چتری [niloofar.moghaddas\[at\]gmail.com](mailto:niloofar.moghaddas[at]gmail.com) را قرار دهید.
2. گروه تلگرام: می توانید برای پرسش سوال از مینی پروژه چهارم در گروه، خانم چتری را با شناسه تلگرامی [@GreenGrain](https://t.me/GreenGrain) در پیام خود نام ببرید.

✓ ایمیل دستیار آموزشی: yeganeh.cht@gmail.com

Hidden Markov Model in Machine Learning

A Hidden Markov Model (HMM) is a statistical model used to describe systems with unobservable states that change over time. It is based on the premise that there is an underlying process with hidden states, each associated with observable outcomes. The model defines probabilities for transitioning between hidden states and for emitting observable symbols. HMMs are widely used in various fields, including finance, bioinformatics, and speech recognition, due to their effectiveness in capturing uncertainty and temporal dependencies. They are particularly valuable for modeling dynamic systems and predicting future states based on observed sequences. An HMM comprises two main types of variables: hidden states and observations.



- The hidden states are the underlying variables that generate the observed data, but they are not directly observable.
- The observations are the variables that are measured and observed.

The relationship between the hidden states and the observations is modeled using a probability distribution. The Hidden Markov Model (HMM) uses two sets of probabilities: the transition probabilities and the emission probabilities.

- The transition probabilities describe the probability of transitioning from one hidden state to another.
- The emission probabilities describe the probability of observing an output given a hidden state.

Project Title: Weather Condition Prediction Using Hidden Markov Models (HMM)

This project aims to [predict weather conditions](#) based on [various meteorological features](#) using Hidden Markov Models (HMM).

Dataset:

The dataset consists of hourly weather data, including temperature, humidity, wind speed, visibility, and atmospheric pressure. The project involves data generation, preprocessing, model training, prediction, and evaluation of the model's performance.

	Date	Temperature	Humidity	Wind Speed (km/h)	Weather Condition	Visibility (km)	Pressure (millibars)
0	2023-01-01 00:00:00	17.0	65.0	17.0	Stormy	1.256012	1026.0
1	2023-01-01 01:00:00	28.0	78.0	5.0	NaN	8.588480	1031.0
2	2023-01-01 02:00:00	22.0	76.0	13.0	Sunny	1.779400	1031.0
3	2023-01-01 03:00:00	20.0	45.0	11.0	Snowy	NaN	1046.0
4	2023-01-01 04:00:00	21.0	43.0	6.0	Sunny	8.034191	1037.0

Weather Condition = {'Sunny', 'Windy', 'Cloudy', 'Rainy', 'Stormy', 'Snowy', 'Foggy', 'Frosty', 'Electrical storms', 'Natural disasters'}

What Do You DO?

Step 1: Data Preprocessing

Perform data preprocessing on the dataset. You may select the most appropriate techniques based on your analysis. Consider the following methods:

- Data wrangling
- Data transformation (e.g., normalization, standardization, discretization)
- Data reduction
- Feature selection
- Feature scaling
- Encoding
- Handling missing values

Please explain your rationale for choosing these specific techniques and present your results. It is not necessary to apply all the methods listed; focus on selecting the most effective ones.

Step 1: Implement the Hidden Markov Model (HMM)

There are several types of Hidden Markov Models (HMM), including:

- Discrete HMM
- Continuous HMM
- Inhomogeneous HMM

Please identify the type of problem you are addressing.

To define the HMM model, you may utilize the hmmlearn library. Which type of HMM model do you consider to be the most suitable for your analysis?

Available models

<code>hmm.CategoricalHMM</code>	Hidden Markov Model with categorical (discrete) emissions.
<code>hmm.GaussianHMM</code>	Hidden Markov Model with Gaussian emissions.
<code>hmm.GMMHMM</code>	Hidden Markov Model with Gaussian mixture emissions.
<code>hmm.MultinomialHMM</code>	Hidden Markov Model with multinomial emissions.
<code>hmm.PoissonHMM</code>	Hidden Markov Model with Poisson emissions.
<code>vhmm.VariationalCategoricalHMM</code>	Hidden Markov Model with categorical (discrete) emissions trained using Variational Inference.
<code>vhmm.VariationalGaussianHMM</code>	Hidden Markov Model with Multivariate Gaussian Emissions trained using Variational Inference.

Step 2: Model Training

1. Split the dataset into training and test sets.
2. Fit the model using the training data.

Step 3: Prediction

Predict the hidden states using the test set.

Step 4: Evaluation

Calculate evaluation metrics to assess the model's performance. The following evaluation metrics are available:

- Confusion Matrix
- Accuracy and Precision
- F1 Score
- ROC Curve
- AUC (Area Under the Curve)

- Precision-Recall Curve
- Classification Reports
- Log Loss
- Recall
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R² Score
- Mean Absolute Percentage Error (MAPE)

You are required to select five metrics from this list, calculate them, and generate corresponding plots. Additionally, please provide a rationale for your selections and present your analysis of the results.

Step 1: Visualization

Finally, create visualizations that include the evaluation metrics, feature importance, actual versus predicted conditions, and any other plots that may aid in making informed decisions.

HAVE FUN 