

ຮູບແບບການສອບເສັງວິຊາ Machine Learning

ຄັງວັນທີ 13/7/2022, ເວລາ 90 ນາທີ

ຊື່ ແລະ ນາມສະກຸນ: ທ້າວ ຮິວ່າງ ຊຸ່ງປໍ

ລະຫັດນັກສຶກສາ: 205Q

ຫ້ອງ: 3CW1

I. ພາກທິດສະດີ

ຈົ່ງຕອບຄໍາຖາມດັ່ງຕໍ່ໄປນີ້ໂດຍການເລືອກ ຫຼືໝາຍເອົາຄໍາຕອບທີ່ຖືກຕ້ອງທີ່ສຸດ (ທັງໝົດ 15 ຂໍ້) ຕົວຢ່າງ:

1. Machine Learning ແມ່ນຫຍັງ?

ກ. ການຮຽນຮູ້ຂອງຄອມພິວເຕີ

ຂ. ການຮຽນຮູ້ຂອງຄອມພິວເຕີຜ່ານຊຸດຂໍ້ມູນອາດີດ

ຄ. ການຮຽນຮູ້ຂອງຄອມພິວເຕີຜ່ານຊຸດຂໍ້ມູນອາດີດເພື່ອສະສົມປະສົບການໃນການຄາດຄະເນຂໍ້ມູນອານາຄົດ

ງ. ຂໍ້ ກ, ຂ ແລະ ຄ ຖືກ

2-15

- ທິດສະດີພາບລວມຂອງ Machine Learning

=> ການຮຽນຮູ້ຂອງເຄື່ອງຈັກ (ML) ແມ່ນປະເພດຂອງປັນຍາປະດິດ (AI) ທີ່ຊ່ວຍໃຫ້ແອັບພລິເຄຊັນຊຸ່ມແວມີຄວາມຖືກຕ້ອງຫຼາຍຂຶ້ນໃນການຄາດເດົາຜົນໄດ້ຮັບໂດຍບໍ່ໄດ້ຕັ້ງໂຄງການຢ່າງຈະແຈ້ງເພື່ອເຮັດແນວນັ້ນ. ຂັ້ນຕອນການຮຽນຮູ້ຂອງເຄື່ອງຈັກໃຊ້ຂໍ້ມູນປະຫວັດສາດເປັນການປ້ອນຂໍ້ມູນເພື່ອຄາດຄະເນຄ່າຜົນຜະລິດໃໝ່.

- ປະເພດຂອງ Machine Learning

=> ປະເພດຂອງ Machine learning ມີ 3 ປະເພດ

- ✓ Supervised learning
- ✓ Unsupervised learning
- ✓ Reinforcement learning

- ເຕັກນິກການຈຳແນກຂໍ້ມູນ (Classification) ແລະ ການວິເຄາະ Regression

1) => Regression)

- Regression algorithms ຖືກນຳໃຊ້ຖ້າມີຄວາມສຳພັນລະຫວ່າງຕົວແປ input ແລະຕົວແປຜົນໄດ້ຮັບ.
- ມັນຖືກນຳໃຊ້ສຳລັບການຄາດຄະເນຂອງຕົວແປຢ່າງຕໍ່ເນື່ອງ, ເຊັ່ນ: ການພະຍາກອນອາກາດ, ແນວໂນ້ມຕະຫຼາດ, ແລະອື່ນໆ.
- Regression algorithms ພາຍໃຕ້ SL
 - Linear Regression
 - Regression Trees
 - Non-Linear Regression

- Bayesian Linear Regression

Polynomial Regression

2) => Classification)

- ສຸດການຄິດໄລ່ການຈັດປະເພດແມ່ນໃຊ້ເນື້ອຕົວແປຜົນໄດ້ຮັບເປັນປະເພດ, ຊຶ່ງໝາຍຄວາມວ່າມີສອງປະເພດເຊັ່ນ Yes-No, Male-Female, True-false, ແລະ ອື່ນໆ.
 - Random Forest
 - Decision Trees
 - Logistic Regression
 - Support vector Machines

- There are classification tasks that you may encounter;

- Binary Classification
- Multi-Class Classification
- Multi-Label Classification
- Imbalanced Classification

- ເຕັກນິກການແບ່ງກຸ່ມຂໍ້ມູນ (Clustering) ແລະ ການສ້າງກົດຄວາມສໍາພັນ (Association)

=> (Clustering)

- ການຈັດກຸ່ມເປັນວິທີການຈັດກຸ່ມວັດຖຸອອກເປັນກຸ່ມ, ມີຄວາມຄ້າຍຄືກັນຫຼາຍທີ່ສຸດ ຫຼືບໍ່ມີຄວາມຄ້າຍຄືກັນ.
- ການວິເຄາະກຸ່ມຊອກຫາຄວາມຄ້າຍຄືກັນລະຫວ່າງວັດຖຸຂໍ້ມູນ ແລະຈັດປະເພດພວກມັນຕາມການມີຢູ່ ແລະບໍ່ມີຂອງຄວາມຄ້າຍຄືກັນເຫຼົ່ານັ້ນ.

=> (Association)

- ກົດລະບຽບສະມາຄົມແມ່ນວິທີການຮຽນຮູ້ທີ່ບໍ່ມີການຄວບຄຸມທີ່ຖືກນໍາໃຊ້ເພື່ອຄົ້ນຫາຄວາມສໍາພັນລະຫວ່າງຕົວແປໃນຖານຂໍ້ມູນຂະໜາດໃຫຍ່.
- ມັນກໍານົດຊຸດຂອງລາຍການທີ່ເກີດຂຶ້ນຮ່ວມກັນໃນຊຸດຂໍ້ມູນ.
- ກົດລະບຽບຂອງສະມາຄົມເຮັດໃຫ້ຍຸດທະສາດການຕະຫຼາດມີປະສິດທິພາບຫຼາຍຂຶ້ນ.
- ເຊັ່ນວ່າຜູ້ທີ່ຊື້ລາຍການ X (ສົມມຸດວ່າເຂົ້າຈີ່) ຍັງມີແນວໂນ້ມທີ່ຈະຊື້ລາຍການ Y (ມັນເບີ / ແຈັມ). ຕົວຢ່າງປົກກະຕິຂອງກົດລະບຽບສະມາຄົມແມ່ນການວິເຄາະຕະຫຼາດຕະຫຼາດ.

- Supervised Learning, Unsupervised Learning ແລະ. Reinforcement Learning

=> Supervised Learning ການຮຽນຮູ້ທີ່ມີການຄວບຄຸມເປັນວິທີການຮຽນຮູ້ເຄື່ອງຈັກທີ່ຕົວແບບໄດ້ຮັບການຝຶກອົບຮົມໂດຍນຳໃຊ້ຂໍ້ມູນຕິດສະຫຼາກ. ໃນການຮຽນຮູ້ແບບຄວບຄຸມ, ຕົວແບບຈຳເປັນຕ້ອງຊອກຫາຫນ້າທີ່ສ້າງແຜນທີ່ເພື່ອເຮັດແຜນທີ່ຕົວແບບການປ້ອນຂໍ້ມູນ (X) ກັບຕົວແບບຜົນໄດ້ຮັບ (Y).

How Supervised Learning Works?

ການຮຽນຮູ້ທີ່ມີການຄວບຄຸມການນຳໃຊ້ຊຸດການຝຶກອົບຮົມເພື່ອສອນຕົວແບບເພື່ອໃຫ້ໄດ້ຮັບຜົນຜະລິດທີ່ຕ້ອງການ. ຊຸດຂໍ້ມູນການຝຶກອົບຮົມນີ້ປະກອບມີວັດສະດຸປ້ອນແລະຜົນໄດ້ຮັບທີ່ຖືກຕ້ອງ, ເຊິ່ງຊ່ວຍໃຫ້ຕົວແບບຮຽນຮູ້ໃນໄລຍະເວລາ. ສຸດການຄິດໄລ່ການວັດແທກຄວາມຖືກຕ້ອງຂອງຕົນໂດຍຜ່ານຫນ້າທີ່ສູນເສຍ, ປັບຈົນກ່ວາຄວາມຜິດພາດໄດ້ຖືກຫຼຸດລົງພຽງພໍ.

=> Unsupervised Learning ການຮຽນຮູ້ແບບບໍ່ມີການຄວບຄຸມ, ເຊິ່ງເອີ້ນກັນວ່າການຮຽນຮູ້ເຄື່ອງຈັກແບບບໍ່ມີການຄວບຄຸມ, ໃຊ້ລະບົບການຮຽນຮູ້ເຄື່ອງຈັກເພື່ອວິເຄາະ ແລະຈັດກຸ່ມຂໍ້ມູນທີ່ບໍ່ມີປ້າຍກຳກັບ. ສຸດການຄິດໄລ່ເຫຼົ່ານີ້ຄົ້ນພົບຮູບແບບທີ່ເຊື່ອງໄວ້ ຫຼືການຈັດກຸ່ມຂໍ້ມູນໂດຍບໍ່ຈຳເປັນຕ້ອງມີການແຊກແຊງຂອງມະນຸດ.

how unsupervised learning works?

ການຮຽນຮູ້ແບບບໍ່ມີການຄວບຄຸມເບິ່ງແຍງເຮັດວຽກແນວໃດ. ເວົ້າງ່າຍໆ, ການຮຽນຮູ້ທີ່ບໍ່ມີການເບິ່ງແຍງເຮັດວຽກໂດຍການວິເຄາະຂໍ້ມູນທີ່ບໍ່ໄດ້ຈັດປະເພດ, ບໍ່ມີປ້າຍຊື່ແລະຊອກຫາໂຄງສ້າງທີ່ເຊື່ອງໄວ້ໃນມັນ.

=> Reinforcement Learning ວິທີການຮຽນຮູ້ທີ່ອີງໃສ່ຄຳຕີຊົມ, ໃນເຊິ່ງຕົວແທນການຮຽນຮູ້ໄດ້ຮັບລາງວັນແຕ່ລະການກະທຳທີ່ຖືກຕ້ອງແລະໄດ້ຮັບການລົງໂທດແຕ່ລະການກະທຳຜິດ.

- ຊະນິດຂໍ້ມູນ ແລະ ການກະກຽມຂໍ້ມູນກ່ອນການປະມວນຜົນ

=>

- ຂັ້ນຕອນວິທີ (Algorithm): Random Forest, Decision Trees, Logistic Regression, NB, KNN, SVM, K-mean...

=> Random Forest

=> Decision Trees

=> Logistic Regression

=>, NB, KNN

=> SVM, K-mean..

- Simple Linear Regression, Multiple Linear Regression ແລະ Polynomial Regression

=> Simple Linear Regression ຖ້າຕົວແບບອິດສະລະອັນດຽວຖືກໃຊ້ເພື່ອຄາດຄະເນຄ່າຂອງຕົວແປທີ່ຂຶ້ນກັບຕົວເລກ, ຫຼັງຈາກນັ້ນ ສຸດການຄິດໄລ່ແບບ Linear Regression ດັ່ງກ່າວເອີ້ນວ່າ Simple Linear Regression.

$$f(x) = M + cx$$

=> Multiple Linear Regression ຖ້າຕົວແປອິດສະລະຫຼາຍກວ່າໜຶ່ງອັນຖືກໃຊ້ເພື່ອຄາດເດົາຄ່າຂອງຕົວແປທີ່ຂຶ້ນກັບຕົວເລກ, ຫຼັງຈາກນັ້ນ ສຸດການຄິດໄລ່ເສັ້ນຊື່ດັ່ງກ່າວເອີ້ນວ່າ Multiple Linear Regression.

$$f(x_1, x_2, \dots) = M + b_1x_1 + b_2x_2, \dots$$

=> Polynomial Regression or PR is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial.

- ຕົວປະເມີນປະສິດຕິພາບ (Metrics) ຂອງການວິເຄາະ Regression

=>

- Dimensionality Reduction)

=>

II. ພາກຄິດໄລ່ (3 ຂໍ້)

1. Simple Linear Regression

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
dataset = pd.read_csv('Salary_Data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/3, random_state = 0)
```

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

```
y_pred = regressor.predict(X_test)
```

```
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Training set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

```
plt.scatter(X_test, y_test, color = 'red')
```

```
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Test set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

2. ຕົວປະເມີນປະສິດທິພາບ (Metrics) ຂອງ Classification: Confusion Matrix

3. ການລຽງລຳດັບຄຳສັ່ງພາສາ Python ໃນຕາຕະລາງໃຫ້ສຳເລັດ ແລະ ຖືກຕ້ອງພ້ອມອະທິບາຍວ່າຄຳສັ່ງດັ່ງກ່າວໃຊ້ເພື່ອເຮັດຫຍັງ, ເພື່ອໃຫ້ປະສົບຜົນສຳເລັດໃນການປະມວນຜົນ.

ຕົວຢ່າງ:

ຕາຕະລາງທີ ລຳດັບຄຳສັ່ງພາສາ Python ພ້ອມຄຳອະທິບາຍ.

ຄຳສັ່ງພາສາ Python	ໃຊ້ເພື່ອ	ລຳດັບ
import numpy as np import matplotlib.pyplot as plt import pandas as pd	ນີ້ແມ່ນພວກ library ທີ່ Import ເຂົ້າມາຊ່ວຍໃນການຈັດການຂໍ້ມູນເພື່ອໃຫ້ເຮັດວຽກໄດ້ງ່າຍ	

dataset = pd.read_csv('Wine.csv') X = dataset.iloc[:, :-1].values y = dataset.iloc[:, -1].values	ໃນປັດຈຸບັນ, ເພື່ອຄາດຄະເນວ່າຜູ້ໃຊ້ຈະຊື້ຜະລິດຕະພັນຫຼືບໍ່, ຄົນເຮົາຕ້ອງຊອກຫາຄວາມສຳພັນລະຫວ່າງອາຍຸແລະເງິນເດືອນທີ່ຄາດຄະເນ. ທີ່ນີ້ User ID ແລະບົດບາດຍິງຊາຍບໍ່ແມ່ນປັດໃຈສຳຄັນສຳລັບການຊອກຫານີ້.	
from sklearn.model_selection import train_test_split X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)	ແຍກຂຸດຂໍ້ມູນເພື່ອຝຶກ ແລະທົດສອບ. 75% ຂອງຂໍ້ມູນແມ່ນໃຊ້ສຳລັບການຝຶກອົບຮົມຕົວແບບແລະ 25% ຂອງມັນຖືກໃຊ້ເພື່ອທົດສອບການປະຕິບັດຂອງຕົວແບບຂອງພວກເຮົາ.	1
from sklearn.preprocessing import StandardScaler sc = StandardScaler() X = sc.fit_transform(X)	ແມ່ນວ່າມັນຈະປ່ຽນຂໍ້ມູນຂອງທ່ານເຊັ່ນວ່າການແຈກຢາຍຂອງມັນຈະມີຄ່າສະເລ່ຍ 0 ແລະມາດຕະຖານ deviation ຂອງ 1. ດຽວນີ້, ມັນເປັນສິ່ງ ສຳ ຄັນຫຼາຍທີ່ຈະປະຕິບັດ	

	<p>ການຂະ ໜາດ ຄຸນສົມບັດຢູ່ທີ່ນີ້ເພາະວ່າຄ່າຂອງ ອາຍຸແລະຄ່າຈ້າງທີ່ຄາດຄະເນຢູ່ໃນຂອບເຂດທີ່ ແຕກຕ່າງກັນ. ຖ້າພວກເຮົາບໍ່ປັບຂະໜາດ ຄຸນສົມບັດ, ຄຸນສົມບັດເງິນເດືອນທີ່ຄາດຄະເນ ຈະຄອບງຳຄຸນສົມບັດຂອງອາຍຸໃນເວລາທີ່ຕົວ ແບບຊອກຫາເພື່ອນບ້ານທີ່ໃກ້ທີ່ສຸດກັບຈຸດຂໍ້ ມູນໃນພື້ນທີ່ຂໍ້ມູນ.</p>	
<pre>from sklearn.linear_model import LogisticRegression classifier = LogisticRegression(random_state = 0) classifier.fit(X_train, y_train)</pre>	<p>ສຸດທ້າຍນີ້ ເຮົາກໍາລັງຝຶກ ໂມເດລ Logistic Regression</p>	
<pre>y_pred = classifier.predict(X_test)</pre>	<p>หลังจากฝึกโมเดลแล้ว ก็ถึงเวลาใช้ แบบจำลองนี้ในการคาดคะเนข้อมูลการ ทดสอบ</p>	