Basic Mobile Lab 2 (521200-1)

Sehwan (Paul) Kim, Ph.D

Associate Professor, Dept. Biomedical Engineering, School of Medicine, Dankook University, Korea
Vice President, Beckman Laser Institute Korea, Dankook University, Korea
Senior Specialist, Beckman Laser Institute, University of California, Irvine, USA

Contents of this week

Basic Terminologies of Machine Learning

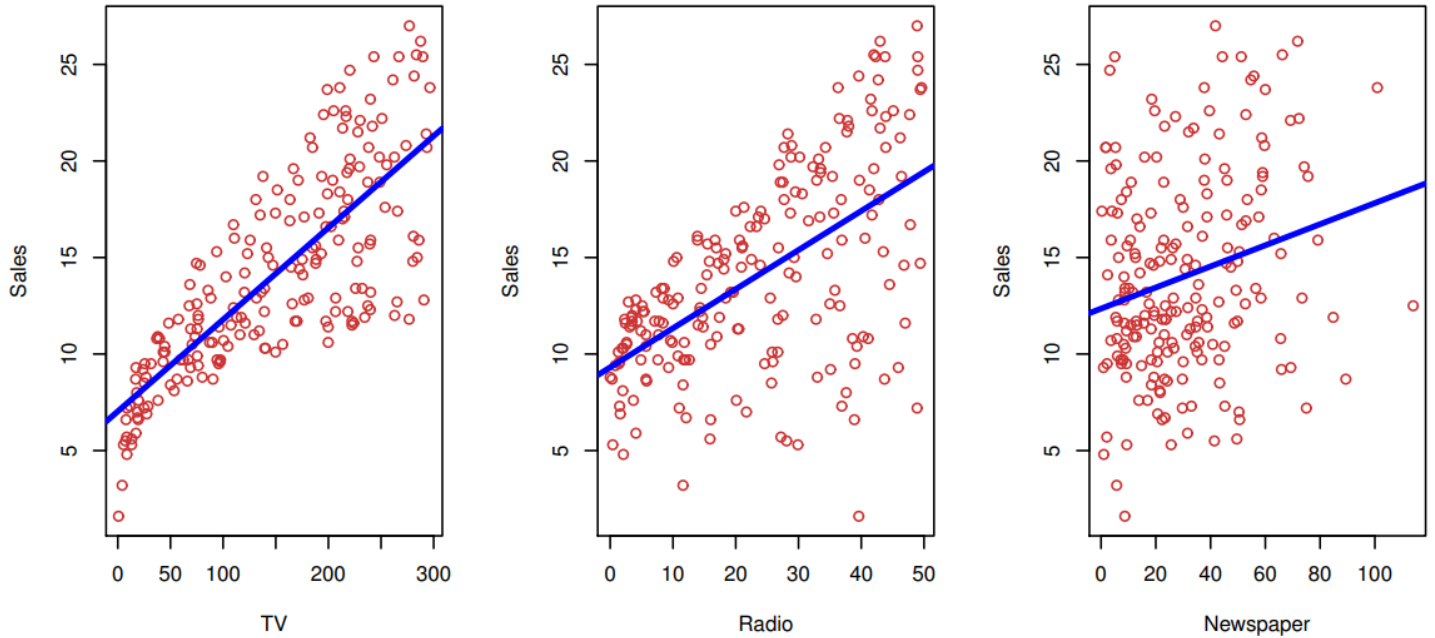
Scikit Learn

1. Why to use it
2. How to use it (lab)
3. Becoming a ML user (slide & lab)

Basic Terminologies of Machine Learning

Let’s use ‘Advertising’ data set as an example.

Dataset consists of the **sales** of the product, along with advertising budgets for three diffent media, : **TV**, **radio**, and **newspaper**.



reference : ISLR

Basic Terminologies of Machine Learning

Terminologies!

Let's assume that you want to predict or explain 'sales',
using budgets of 'TV', 'Radio', and 'Newspaper'

'TV', 'Radio', 'Newspaper' : **Input variable**, typically denoted using x_1, x_2, x_3
Inputs can have various different names :
predictors, independent variables, features, or just variables

'Sales' : **Output variable**, typically denoted using y
output can have various different names :
response variable, dependent variable, target

Terminologies : supervised learning

When we do supervised learning, we assume that there is some relationship between **y** and $\mathbf{X} = (x_1, x_2, x_3, .. , x_p)$

$$\mathbf{y} = f(\mathbf{X}) + \varepsilon$$

f : fixed, unknown function of **X**, relationship between **y** and **X**. systematic information that **X** provides about **y**.

ε : random error term, noise of the data itself, irreducible. cannot be predicted using **X**.

Terminologies : supervised learning

When we want to make a prediction or inference,
we are making $\hat{f}$ **as an estimate for** f

$$y = f(X) + \varepsilon$$

$$\hat{y} = \hat{f}(X)$$

$\hat{f}$: an estimated relationship (a ML model you choose)
$\hat{y}$: prediction for y

Basic Terminologies of Machine Learning

Terminologies : supervised learning

$$\begin{aligned} \mathrm{E}(Y - \hat{Y})^2 &= \mathrm{E}[f(X) + \epsilon - \hat{f}(X)]^2 \\ &= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{Reducible}} + \underbrace{\mathrm{Var}(\epsilon)}_{\text{Irreducible}} \end{aligned}$$

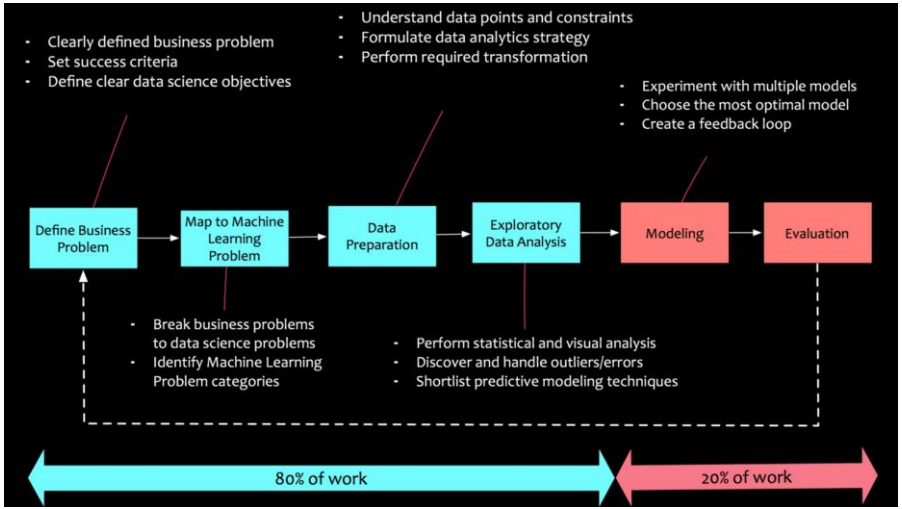
You will get your trained model, after minimizing **error**

Reducible Error :
- Error between true unknown relationship between estimated relationship.

Irreducible Error :
- Error in the data itself, **ε** cannot be predicted using **X**.

Scikit learn : why to use it?

Scikit learn : why to use it?

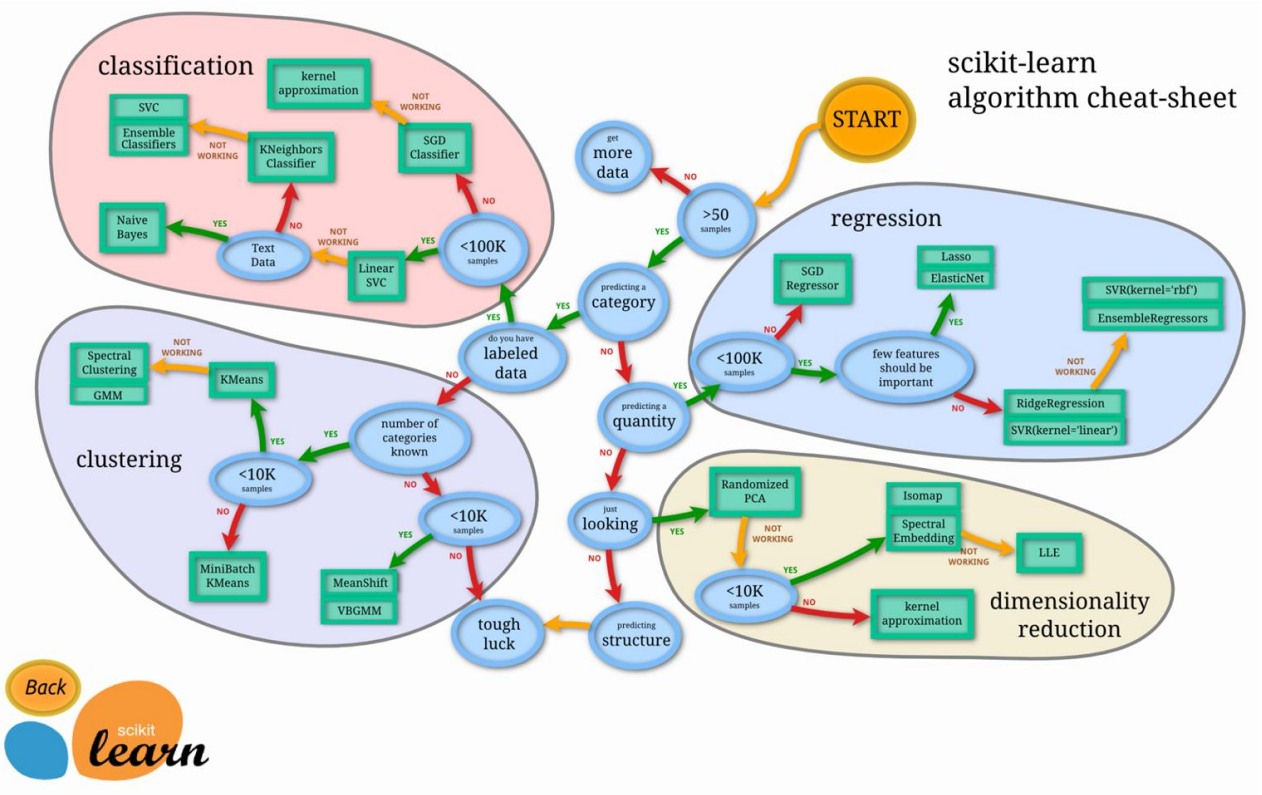


In the real machine learning Project

- There are lots of things to do, other than machine learning itself.
- Implementing a ML algorithm whenever we need is not efficient way.
- Fast prototyping is as Important as theoretical understanding of ML

Scikit learn : why to use it?

Sklearn already contains lots of ML algorithms



Scikit learn : why to use it?

Sklearn is really easy to use

```
[5]  1  # 1. Import what model you want.
     2  from sklearn.ensemble import RandomForestRegressor
     3
     4
     5  # 2. Declare your model.
     6  rf = RandomForestRegressor(  )
     7
     8
     9  # 3. Fit your model.
    10  rf.fit(x_in_sample, y_in_sample)
    11
    12
    13  # 4. predict using your fitted model.
    14  y_pred = rf.predict(x_new_observed)
```

These 4 lines of code is just enough to start

- You can use ML algorithms before learning ML.
- You can do more after learning ML, in addition to these 4 lines of code.

Scikit learn : why to use it?

Sklearn is really easy to use

```
[5]  1  # 1. Import what model you want.
     2  from sklearn.ensemble import RandomForestRegressor
     3
     4
     5  # 2. Declare your model.
     6  rf = RandomForestRegressor(  )
     7
     8
     9  # 3. Fit your model.
    10  rf.fit(x_in_sample, y_in_sample)
    11
    12
    13  # 4. predict using your fitted model.
    14  y_pred = rf.predict(x_new_observed)
```

Someday...

- Depending on your task, you will be able to choose the right algorithm.
- Once you understand how the algorithm works, you will be able to set the appropriate hyperparameters.

Scikit learn : why to use it?

APIs of many ML libraries mirror sklearn's API

```
######cuML######

from cuml import
RandomForestClassifier as cuRF


# cuml Random Forest params
cu_rf_params = {
    'n_estimators': 25,
    'max_depth': 13,
     'n_bins': 15,
     'n_streams': 8 }

cu_rf = cuRF(**cu_rf_params)
cu_rf.fit(X_train, y_train)

print("cuml RF Accuracy Score: "
accuracy_score(
cu_rf.predict(X_test), y_test))
```

```
######Sklearn######

from sklearn.ensemble import
RandomForestClassifier as sklRF
import multiprocessing as mp


#sklearn Random Forest params
skl_rf_params = {
    'n_estimators': 25,
    'max_depth': 13,

    'n_jobs': mp.cpu_count() }

skl_rf = sklRF(**skl_rf_params)
skl_rf.fit(X_train, y_train)

print("sklearn RF Accuracy Score: "
accuracy_score(
skl_rf.predict(X_test), y_test))
```

These 4 lines of code is just enough again

1. Import what model you want.
2. Declare your model.
3. Fit your model.
4. Predict using your fitted model.

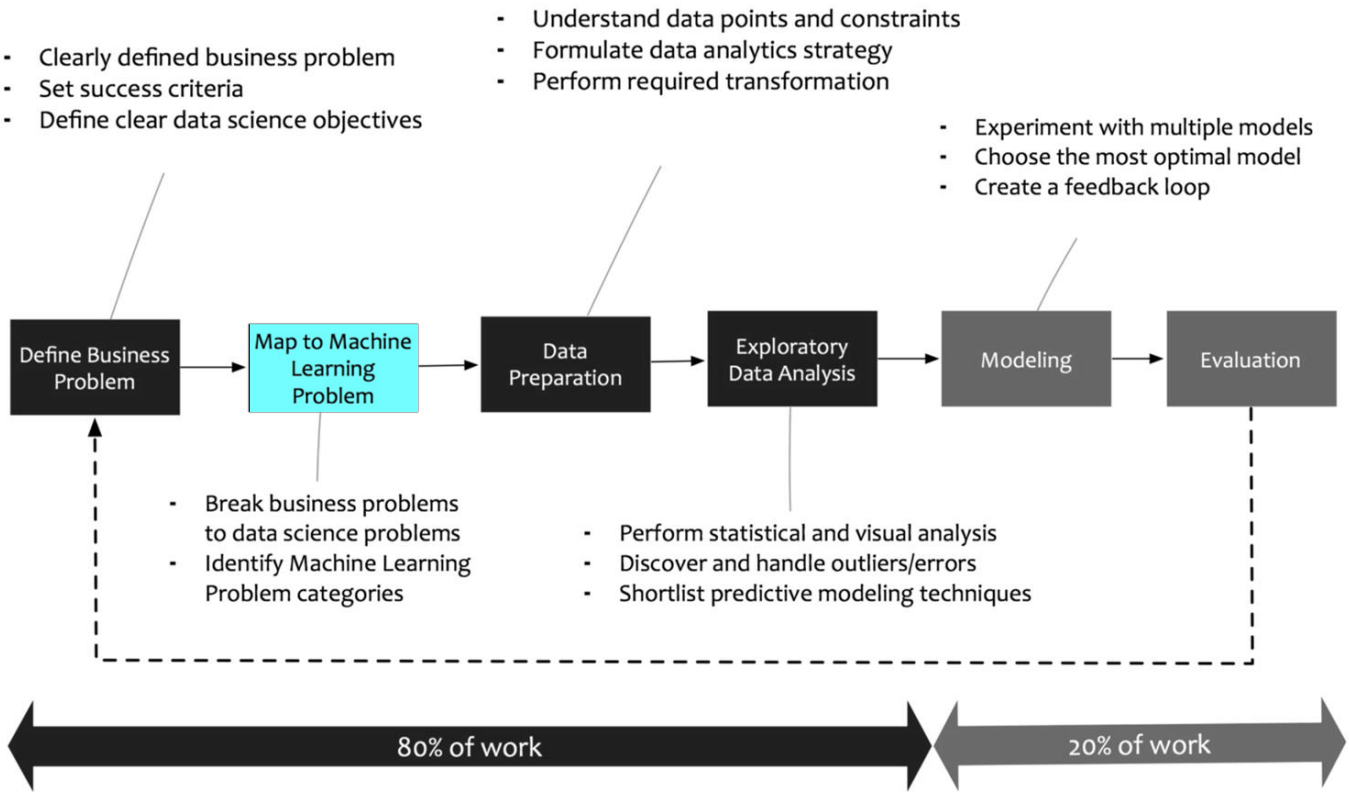
note : rapids.ai(that includes cuML) is an open source project, supported by NVIDIA

Summary

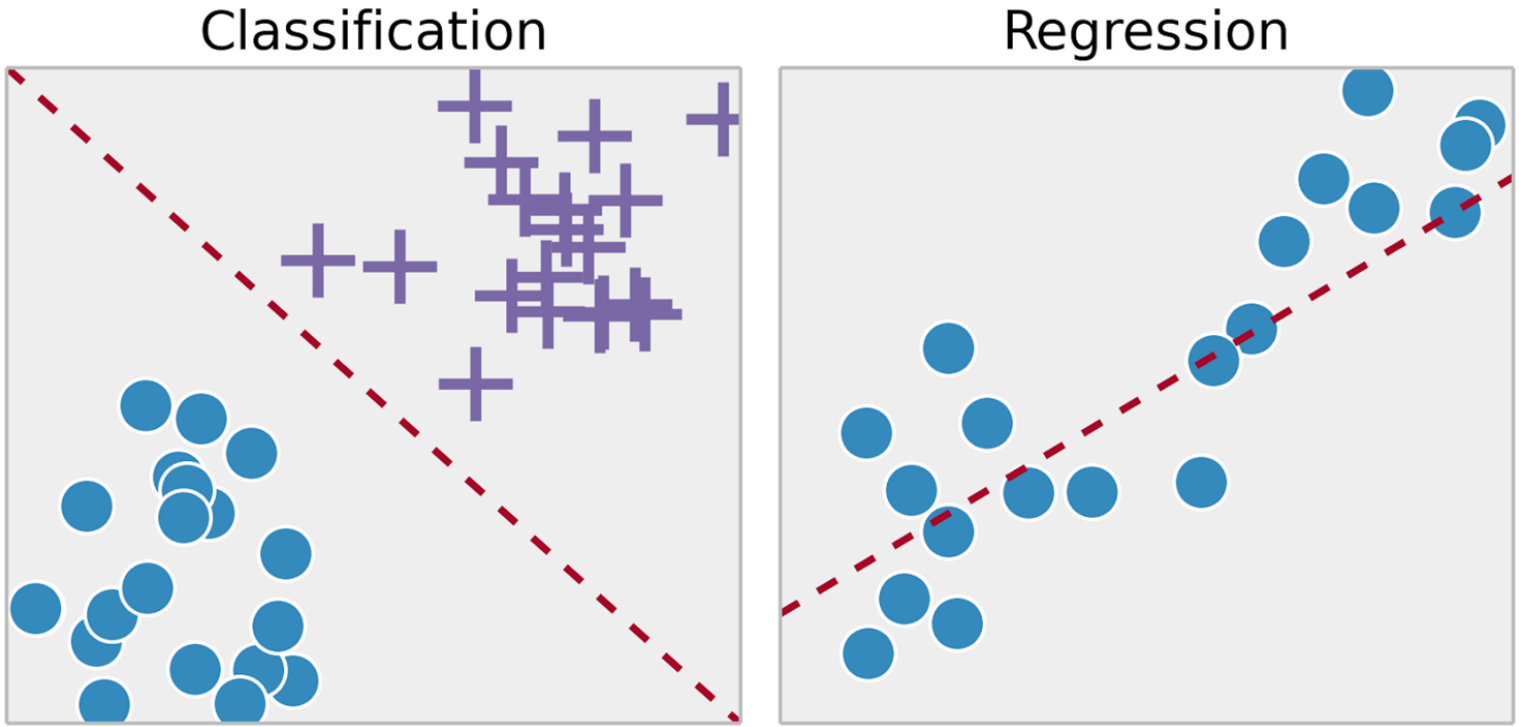
You can be a ML user first with sklearn.

- Many ML algorithms are prepared to you.
  → You can focus entire whole ML process.
- Really easy to use.
  → Familiarity with ML will lead you to become an ML engineer.
- The APIs of many other ML libraries is very simliar to sklearn’s API.
  → You can learn other ML libraries quickly.

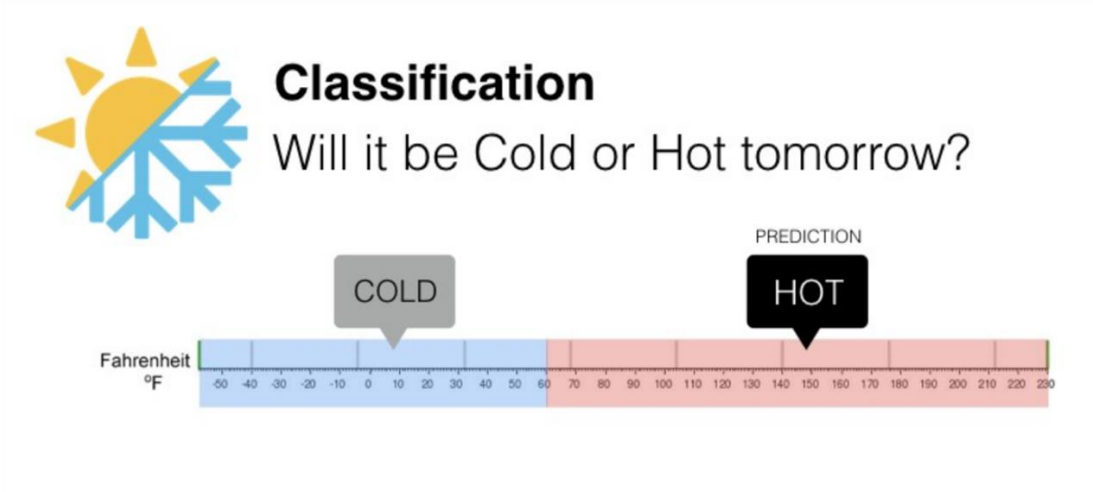
Becoming a ML user



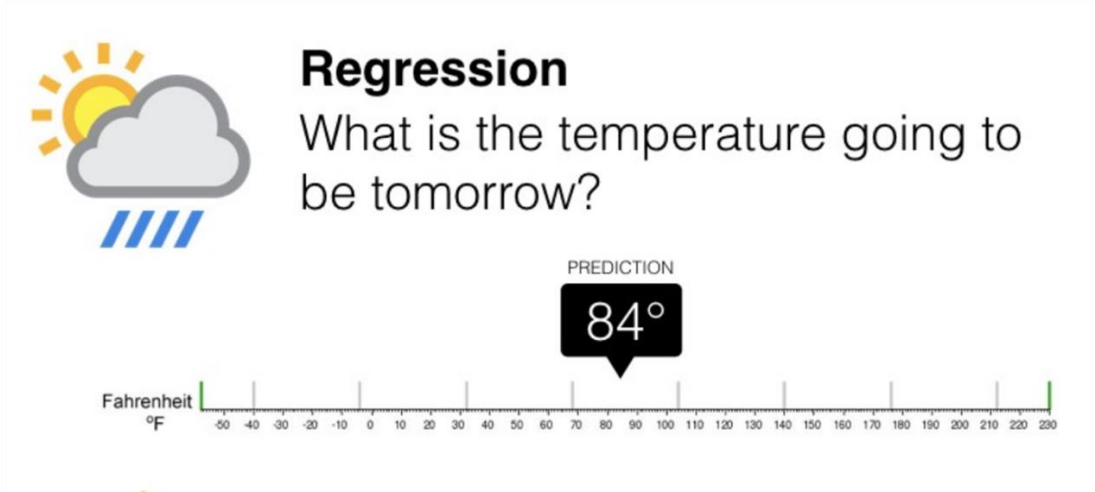
Map to Machine Learning Problem in basic level



Classification or Regression



Classification or Regression



Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		
How much will be closing price ?		
The closing price will be higher than now or not?		
a Cat or a Dog or a Person ?		
Rain or Not tomorrow ?		
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?		
The closing price will be higher than now or not?		
a Cat or a Dog or a Person ?		
Rain or Not tomorrow ?		
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?	**Yes**	
The closing price will be higher than now or not?		
a Cat or a Dog or a Person ?		
Rain or Not tomorrow ?		
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?	**Yes**	
The closing price will be higher than now or not?		**Yes**
a Cat or a Dog or a Person ?		
Rain or Not tomorrow ?		
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?	**Yes**	
The closing price will be higher than now or not?		**Yes**
a Cat or a Dog or a Person ?		**Yes**
Rain or Not tomorrow ?		
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?	**Yes**	
The closing price will be higher than now or not?		**Yes**
a Cat or a Dog or a Person ?		**Yes**
Rain or Not tomorrow ?		**Yes**
The probability of rain tomorrow ?		

Quiz & Summary

You should be able to distinguish problems.

	Regression	Classification
SPAM or Not ?		**Yes**
How much will be closing price ?	**Yes**	
The closing price will be higher than now or not?		**Yes**
a Cat or a Dog or a Person ?		**Yes**
Rain or Not tomorrow ?		**Yes**
The probability of rain tomorrow ?	**Yes**	

Quiz & Summary

Yes, Confusing.

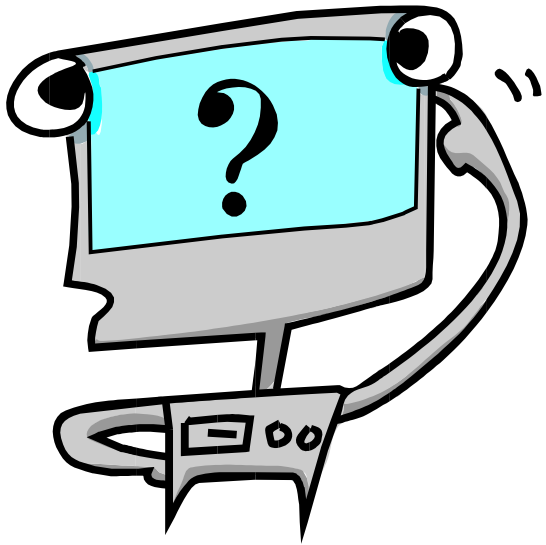
	Regression	Classification
SPAM or Not ?		Yes
How much will be closing price ?	Yes	
The closing price will be higher than now or not?		**Yes**
a Cat or a Dog or a Person ?		Yes
Rain or Not tomorrow ?		Yes
The probability of rain tomorrow ?	**Yes**	

Quiz & Summary

	Regression	Classification
The closing price will be higher than now or not?		**Yes**
The probability of rain tomorrow ?	**Yes**	

At first, Just focus the final result.

Someday, you will solve ‘classification’ problem,
using ‘regression’ techniques.




THANK YOU