

01 Practice

1. Consider this subset of `mtcars`:

	car	mpg	cyl	hp	wt	vs	am
	Mazda RX4	21.0	6	110	2.620	0	1
	Mazda RX4 Wag	21.0	6	110	2.875	0	1
	Datsun 710	22.8	4	93	2.320	1	1
	Hornet 4 Drive	21.4	6	110	3.215	1	0
	Hornet Sportabout	18.7	8	175	3.440	0	0

Suppose we want to predict `vs` (where 0 = V engine, 1 = straight) from `cyl`, `hp`, and `wt` (in thousands of pounds). Then:

(a) $N =$ **ANSWER: 5**, (b) $D =$ **ANSWER: 3**, (c) $\mathbf{x}_2$ (boldface)
 $=$ **ANSWER: (6, 110, 2.875)** ,

(d) $y_2 =$ **ANSWER: 0**, (e) $x_2^{(1)} =$ **ANSWER: 110**

2. Using code like the code presented in lecture, train an SVM to predict `vs` (after changing each 0 to -1) from `cyl`, `hp`, and `wt`. Here is code to get started:

```
import numpy as np
import pandas as pd
from sklearn import svm
df = pd.read_csv('http://pages.stat.wisc.edu/~jgillett/451/data/mtcars.csv', index_col=0)
df = df.iloc[0:5, :] # use only first 5 rows
X = df[['cyl', 'hp', 'wt']] # 2D object
y = df.vs # 1D object
y[y == 0] = -1 # convert y from 0s and 1s to -1s and 1s
```

Confirm that the SVM has decision boundary $\mathbf{w}\mathbf{x} + b = 0$, where $\mathbf{w} \approx (-0.04, -0.3, 5.9)$ and $b \approx 15.9$.

You should be able to answer the following question with or without a computer.

- (a) How does the SVM classify a car having `cyl` = 6, `hp` = 100, and `wt` = 2? $\hat{y} =$
☐ -1 (V) **ANSWER: ☒**, ☐ 1 (straight)
- (b) How many errors does the classifier make on the training data?
☐ 0 **ANSWER: ☒**, ☐ 1, ☐ 2, ☐ 3, ☐ 4, ☐ 5
- (c) Out of the five training examples, how many are ($\approx$) support vectors? Please use these more precise versions of the model numbers: $\mathbf{w} = (-0.036, -0.3058, 5.901)$ and $b = 15.8917$. Hint: Do this in parallel by assigning one student to each of the examples.
☐ 0, ☐ 1, ☐ 2, ☐ **ANSWER: ☒ 3**, ☐ 4, ☐ 5
- (d) The code line `y[y == 0] = -1` (above) says “Change the `y` values that were equal to 0 to -1 .” What is the purpose of this line?

☐ It is easier to separate -1 s from 1 s than 0 s from 1 s.

☐ **ANSWER:** ☒ The SVM is a binary classifier whose mathematical description uses $y \in \{-1, 1\}$.

☐ The line is to distract students who just want to learn basic ML.