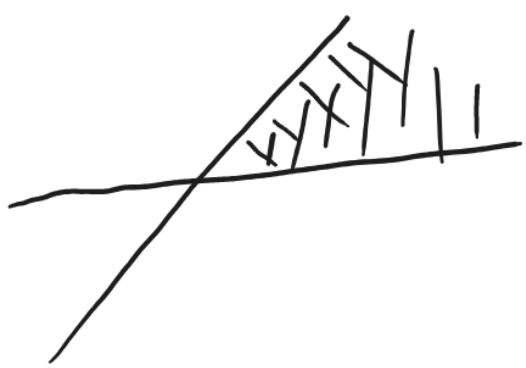
# DSCI 552 Assignment 1

Due January 25, 2023

1. Alpaydin 4th edition, Chapter 2, Exercise 1. What is the VC dimension of a circle?
2. Alpaydin 4th edition, Chapter 2, Exercise 2. Why would setting $m = N$ (where $N$ is the number of positive instances) be a bad idea.
3. Alpaydin 4th edition, Chapter 2, Exercise #6
4. Alpaydin 4th edition, Chapter 2, Exercise #7
5. Alpaydin 4th edition, Chapter 2, Exercise #9
6. Bonus: Let a wedge shape be defined as the intersection of two half-spaces, see the image below. Show the VC dimension of a wedge is 5. Hint: place the five points equidistant on a circle.



From Alpaydin Chapter 2:

1. Let us say our hypothesis class is a circle instead of a rectangle. What are the parameters? How can the parameters of a circle hypothesis be calculated in such a case? What if it is an ellipse? Why does it make more sense to use an ellipse instead of a circle?

   SOLUTION: In the case of a circle, the parameters are the center and the radius (see figure 2.11). We then need to find $S$ and $G$ where $S$ is the tightest circle that includes all the positive examples and $G$ is the largest circle that includes all the positive examples and no negative example; any circle between them is a consistent hypothesis.

   It makes more sense to use an ellipse because the two axes need not have the same scale and an ellipse has two separate parameters for the widths in the two axes rather than a single radius. Actually, price and engine power are positively correlated; the price of a car tends to increase as its engine power increases, and hence it makes more sense to use an oblique ellipse—we will see such models in chapter 5.

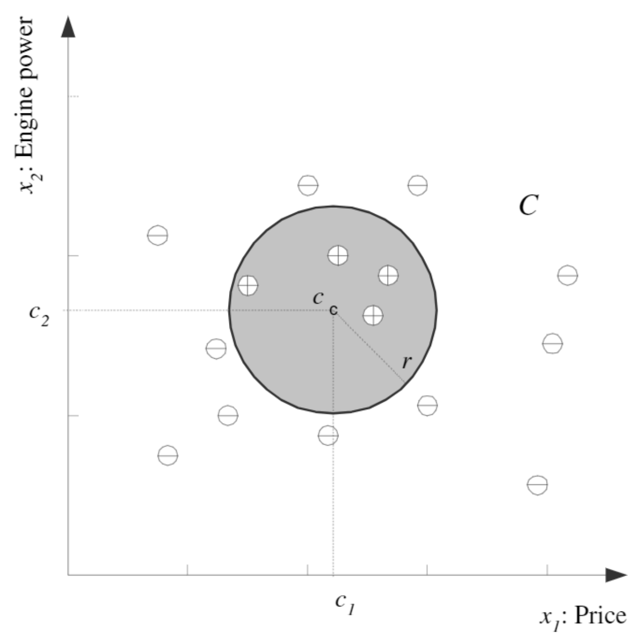



**Figure 2.11** Hypothesis class is a circle with two parameters, the coordinates of its center and its radius.

2. Imagine our hypothesis is not one rectangle but a union of two (or $m > 1$) rectangles. What is the advantage of such a hypothesis class? Show that any class can be represented by such a hypothesis class with large enough $m$.

   SOLUTION: In the case when there is a single rectangle, all the positive instances should form one single group; with two rectangles, for example (see figure 2.12), the positive instances can form two, possibly disjoint clusters in the input space. Note that each rectangle corresponds to a conjunction on the two input attributes, and having multiple rectangles corresponds to a disjunction. Any logical formula can be written as a disjunction of conjunctions.

In the worst case ($m = N$), we have a separate rectangle for each positive instance.

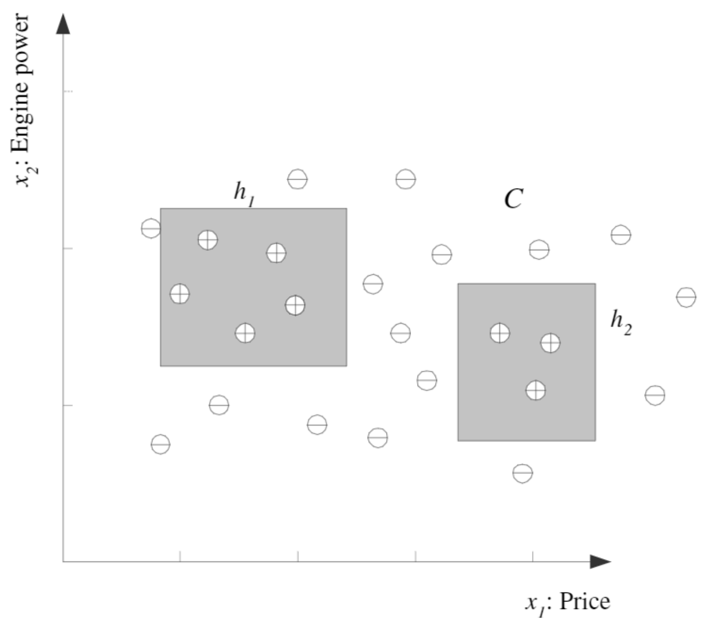



**Figure 2.12** Hypothesis class is a union of two rectangles.

6. In equation 2.13, we summed up the squares of the differences between the actual value and the estimated value. This error function is the one most frequently used, but it is one of several possible error functions. Because it sums up the squares of the differences, it is not robust to outliers. What would be a better error function to implement *robust regression*?

7. Derive equation 2.17.

8. How can we estimate $w_2, w_1, w_0$ for the quadratic model of equation 2.18 ?

9. Assume our hypothesis class is the set of lines, and we use a line to separate the positive and negative examples, instead of bounding the positive examples as in a rectangle, leaving the negatives outside (see figure 2.13). Show that the VC dimension of a line is 3.

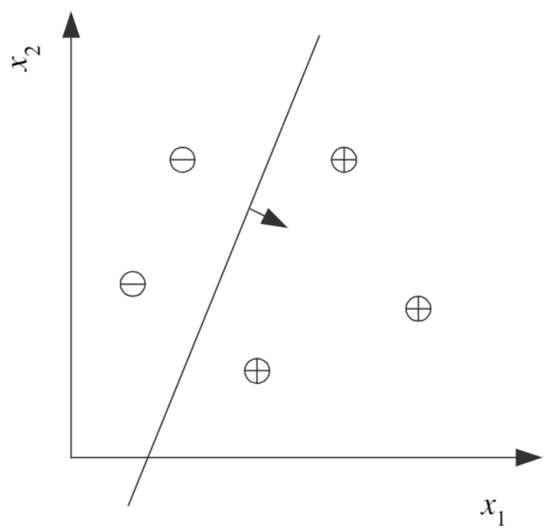



**Figure 2.13** A line separating positive and negative instances.

Alpydin, Equation 2.17

Let us now go back to our example in section 1.2.3 where we estimated the price of a used car. There we used a single input linear model

$$g(x) = w_1 x + w_0 \tag{2.15}$$

where $w_1$ and $w_0$ are the parameters to learn from data. The $w_1$ and $w_0$ values should minimize

$$E(w_1, w_0|\mathcal{X}) = \frac{1}{N}\sum_{t=1}^{N}[r^t - (w_1 x^t + w_0)]^2 \tag{2.16}$$

Its minimum point can be calculated by taking the partial derivatives of $E$ with respect to $w_1$ and $w_0$, setting them equal to 0, and solving for the two unknowns:

$$\begin{aligned} w_1 &= \frac{\sum_t x^t r^t - \overline{x}\,\overline{r}N}{\sum_t (x^t)^2 - N\overline{x}^2} \\ w_0 &= \overline{r} - w_1\overline{x} \end{aligned} \tag{2.17}$$

where $\overline{x} = \sum_t x^t/N$ and $\overline{r} = \sum_t r^t/N$. The line found is shown in figure 1.2.