

Espero que todos vieron y estudiaron el material en los videos (Lecture1.1, 1.2, 1.3) donde Andrew Ng explica los conceptos básicos de aprendizaje de máquina.

¿Cómo la máquina aprende? Le damos datos, también se llaman ejemplos. Cada ejemplo es un vector numérico que representa el dato correspondiente. Para imágenes, el vector va a contener los valores de intensidades de colores de píxeles, para nosotros son características que estudiamos: frecuencia bruta, frecuencia relativa (probabilidad), tf-idf, etc. de palabras. Los vectores de ejemplos se representan como  $x^{(i)}$  donde  $i$  es el número del ejemplo. Los elementos (features) del vector también se representan por  $x$ , pero por  $x$  con subíndice, por ejemplo,  $x_j$  donde  $j$  es el número del elemento del vector.

Entonces cada dato es el vector de este tipo:

$x^{(i)} = (x_1, x_2, \dots, x_j, \dots, x_n)$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq n$ ,  $m$  es el número de ejemplos y  $n$  es el número de características en un vector (longitud del vector).

Cada vector  $x^{(i)}$  está asociado con un valor que se representa por  $y^{(i)}$ . Entonces el conjunto de datos del tamaño  $m$  va a contener los pares de tipo  $(x^{(i)}, y^{(i)})$ .

La idea de aprendizaje es enseñar a la máquina estos ejemplos (pares de tipo  $(x^{(i)}, y^{(i)})$ ), para que ella aprende predecir el valor de  $y^{(i)}$  a partir del vector  $x^{(i)}$ . ¿Como precisamente se realiza este aprendizaje? La máquina toma un vector  $x^{(i)}$  y calcula el valor de  $y$  que le corresponde a  $x^{(i)}$  usando una cierta función o proceso. Este valor calculado de  $y$  se llama el valor estimado de  $y$  y se representa por  $\hat{y}^{(i)}$ . Este valor  $\hat{y}^{(i)}$  normalmente es diferente del valor real  $y^{(i)}$ , es decir, hay un error de estimación. El error se puede calcular usando el método de diferencias cuadradas:

$$\text{error de estimación (o predicción)} = (\hat{y}^{(i)} - y^{(i)})^2$$

El error de estimación o predicción también se llama **costo**. Para cada ejemplo se calcula el error de predicción, luego se suman todos los errores y se promedian por el número  $m$  de ejemplos. Este promedio se llama **función de costo** que se denomina por la letra  $J$ :

$$J = \frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$$

La función de costo a veces se llama función de pérdida o función de error.

Ahora lo que hay que hacer es optimizar la función de costo, es decir, minimizar el error de predicción. El proceso de optimización se llama aprendizaje.

Por favor vean los videos Lecture 2.1, 2.2, 2.3 y 2.4, estudian bien su contenido. En estas clases ustedes van a conocer como se calcula el valor estimado de  $\hat{y}^{(i)}$ .