# Gradient Descent

This notebook will illustrate gradient descent.

Consider the mean squared error loss function on a dataset $\mathcal{D} = [(x_1, x_2), \ldots [x_n, y_n)]$ consisting of $n$ datapoints where $y \in \mathbb{R}$ and $x \in \mathbb{R}^d$.

We are trying to learn a linear mapping $\hat{y}_i = w^T x_i$ which minimizes $\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$. So, our loss function is:

$$L(w) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

If we differentiate with respect to some $w_j$:

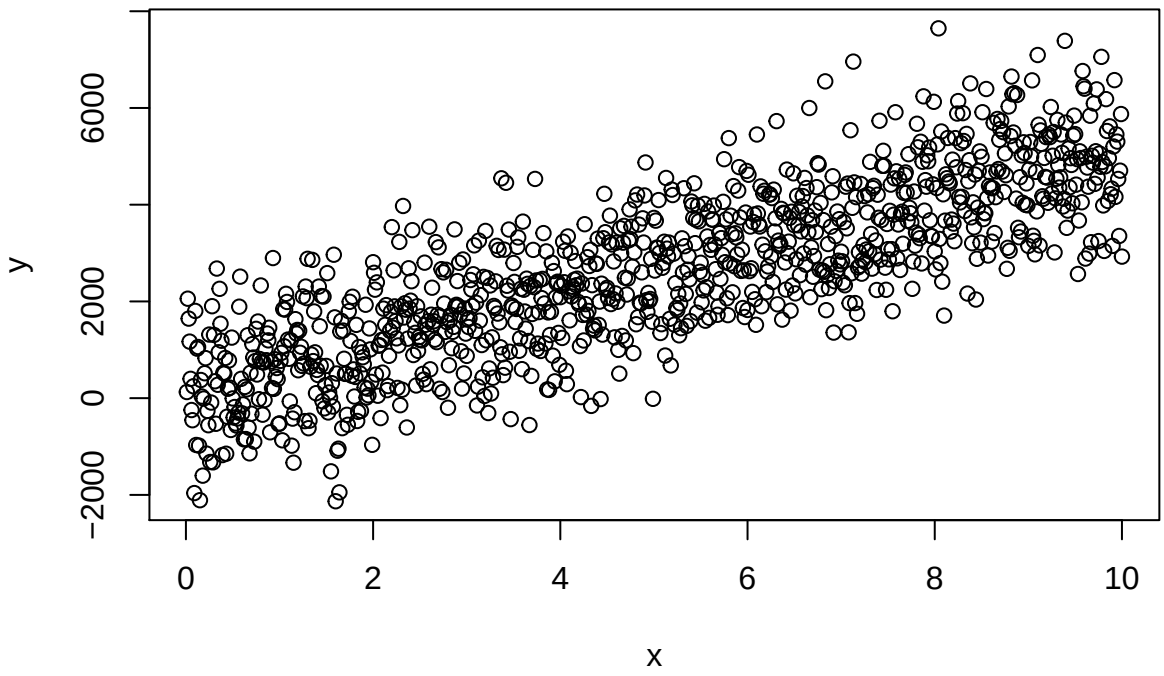
$$\begin{aligned}
\frac{dL_i}{dw_j} &= \frac{d}{dw_j}(w^T x_i - y_i)^2 \\
&= \frac{d}{dw_j}(w^T x_i)^2 + y_i^2 - 2w^T x_i y_i \\
&= 2w_j x_i^j - 2x_i^j y_i
\end{aligned}$$

We've mixed sub and superscripts here but you get the idea.

$$\begin{aligned}
\frac{dL_i}{dw} &= 2x_i w^T x_i - 2x_i y_i \\
\frac{dL}{dw} &= \sum_{i=1}^{n} 2x_i (w^T x_i - y_i)
\end{aligned}$$

Let's generate some data points using some fixed slope and bias, perturbing with Gaussian noise.

```
NUM_SAMPLES = 1000
SLOPE = 500
BIAS = 99

x = seq(0.01,10,by=0.01)
y = seq(0.01,10,by=0.01)

for (i in 1:length(x)) {
  y[i] = SLOPE * x[i] + BIAS + rnorm(1, mean=0, sd=1000)
}

plot(x,y)
```

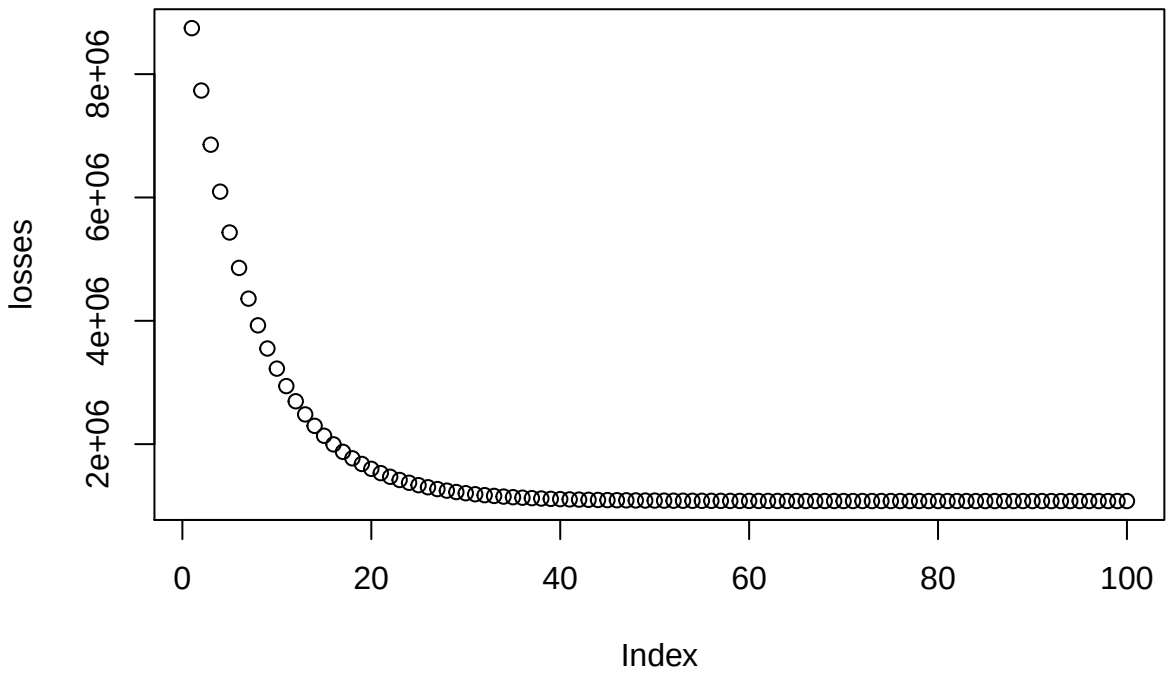



```
step_size = 1e-3

# Gradient of mean squared error.
gradient = function(w) {
  acc = c(0,0)
  for (i in 1:NUM_SAMPLES) {
    xi = c(1, x[i])
    acc = acc + 2 * xi * (sum(w * xi) - y[i])
  }
  return(acc / NUM_SAMPLES)
}

# Average squared error.
loss = function(w) {
  acc = 0
  for (i in 1:NUM_SAMPLES) {
    xi = c(1, x[i])
    acc = acc + (y[i] - sum(xi * w)) ^ 2
  }
  return(acc / NUM_SAMPLES)
}
```

Now, let's run gradient descent.

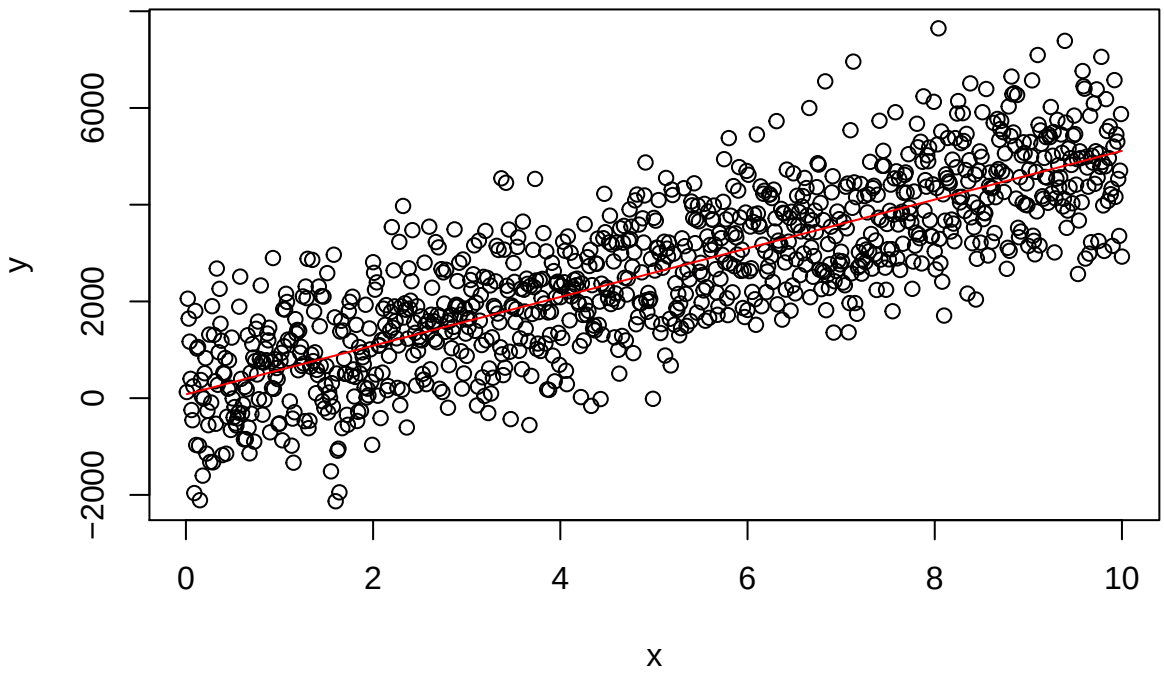
```
w = c(1, 1)
losses = c()
```

```
for (i in 1:100) {
  grad = gradient(w)
  w = w - step_size * grad
  losses[i] = loss(w)
}

plot(losses)
```




Let's plot the line that we find with gradient descent.

```
plot(x, y)
min_y = min(x) * w[2] + w[1]
max_y = max(x) * w[2] + w[1]
lines(c(min(x), max(x)), c(min_y, max_y), col="red")
```


y
6000
2000
0
−2000
0
2
4
6
8
10
x