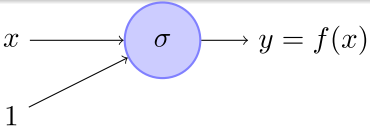
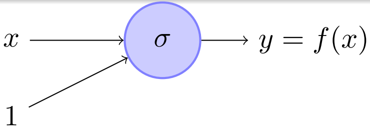
Module 5.1: Learning Parameters : Infeasible (Guess Work)



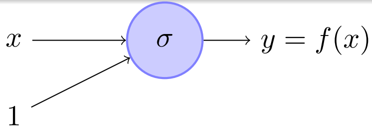
$$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$$



$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$

Input for training

$\{x_i, y_i\}_{i=1}^{N} \rightarrow N$ pairs of (x, y)



$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$

Input for training

$\{x_i, y_i\}_{i=1}^{N} \rightarrow N$ pairs of (x, y)

Training objective

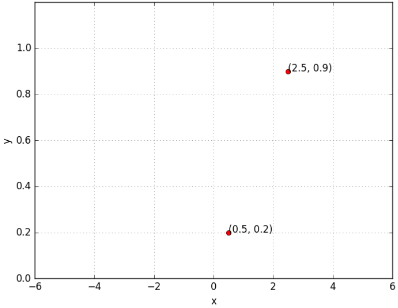
Find w and b such that:

$$\underset{w,b}{\text{minimize}} \, \mathscr{L}(w, b) = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

$$x \longrightarrow \sigma \longrightarrow y = f(x)$$

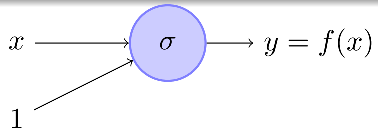
$$1 \nearrow \sigma$$

$$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$$

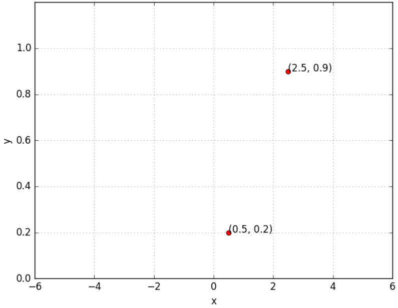


What does it mean to train the network?

- Suppose we train the network with $(x, y) = (0.5, 0.2)$ and $(2.5, 0.9)$

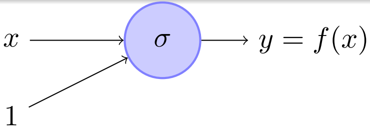


$$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$$

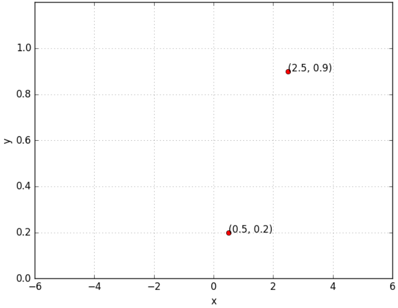


What does it mean to train the network?

- Suppose we train the network with $(x, y) = (0.5, 0.2)$ and $(2.5, 0.9)$
- At the end of training we expect to find w^*, b^* such that:

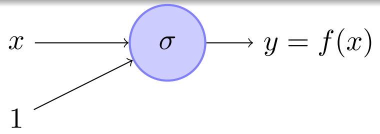


$$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$$

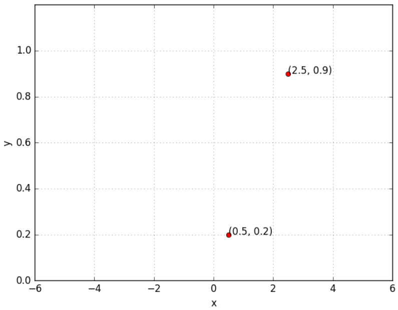


What does it mean to train the network?

- Suppose we train the network with $(x, y) = (0.5, 0.2)$ and $(2.5, 0.9)$
- At the end of training we expect to find w^*, b^* such that:
- $f(0.5) \rightarrow 0.2$ and $f(2.5) \rightarrow 0.9$



$$f(x) = \frac{1}{1+e^{-(w \cdot x+b)}}$$

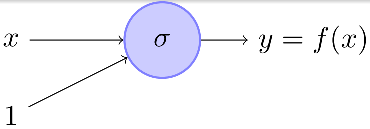


What does it mean to train the network?

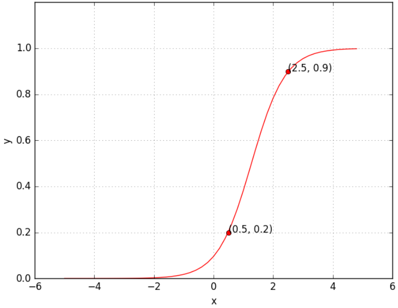
- Suppose we train the network with $(x, y) = (0.5, 0.2)$ and $(2.5, 0.9)$
- At the end of training we expect to find w^*, b^* such that:
- $f(0.5) \rightarrow 0.2$ and $f(2.5) \rightarrow 0.9$

In other words...

- We hope to find a sigmoid function such that $(0.5, 0.2)$ and $(2.5, 0.9)$ lie on this sigmoid



$$f(x) = \frac{1}{1+e^{-(w \cdot x + b)}}$$



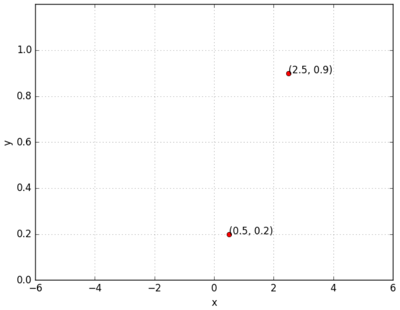
What does it mean to train the network?

- Suppose we train the network with $(x, y) = (0.5, 0.2)$ and $(2.5, 0.9)$
- At the end of training we expect to find w^*, b^* such that:
- $f(0.5) \rightarrow 0.2$ and $f(2.5) \rightarrow 0.9$

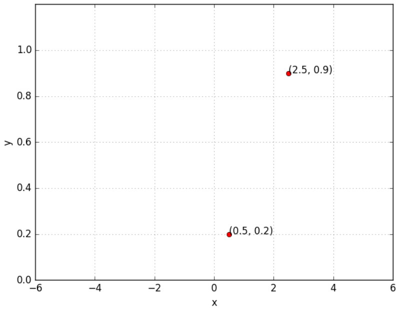
In other words...

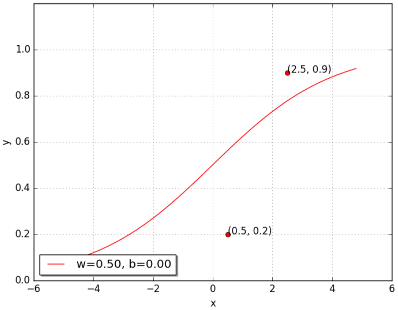
- We hope to find a sigmoid function such that $(0.5, 0.2)$ and $(2.5, 0.9)$ lie on this sigmoid

Let us see this in more detail....

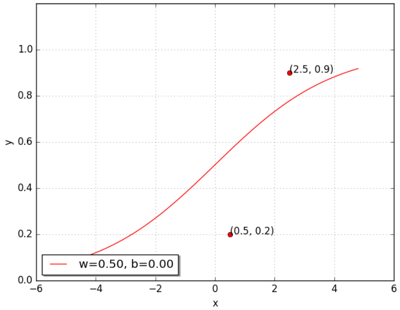

1.0
0.8
0.6
0.4
0.2
0.0
y
−6
−4
−2
0
2
4
6
x
(2.5, 0.9)
(0.5, 0.2)

- Can we try to find such a w^*, b^* manually

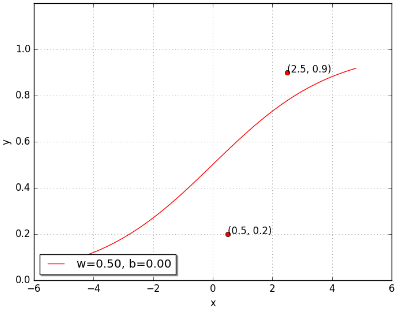




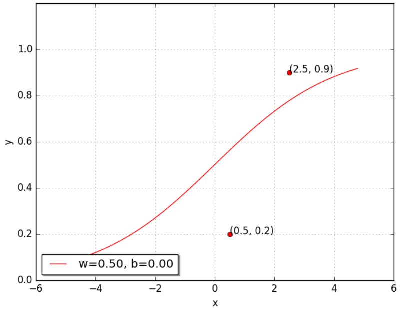
- Can we try to find such a w^*, b^* manually
- Let us try a random guess.. (say, $w = 0.5, b = 0$)



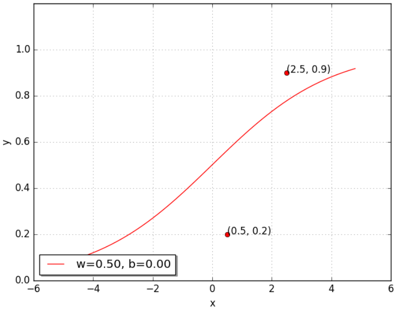
- Can we try to find such a w^*, b^* manually
- Let us try a random guess.. (say, $w = 0.5, b = 0$)
- Clearly not good, but how bad is it ?



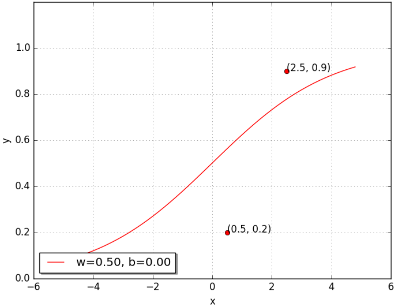
- Can we try to find such a w^*, b^* manually
- Let us try a random guess.. (say, $w = 0.5, b = 0$)
- Clearly not good, but how bad is it ?
- Let us revisit $\mathscr{L}(w, b)$ to see how bad it is ...



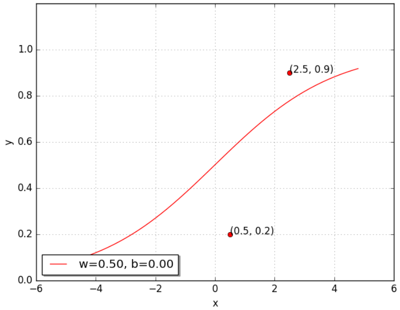
$$\mathscr{L}(w,b) = \frac{1}{2} * \sum_{i=1}^{N} (y_i - f(x_i))^2$$



$$\mathscr{L}(w,b) = \frac{1}{2} * \sum_{i=1}^{N} (y_i - f(x_i))^2$$
$$= \frac{1}{2} * ((y_1 - f(x_1))^2 + (y_2 - f(x_2))^2)$$



$$\mathscr{L}(w,b) = \frac{1}{2} * \sum_{i=1}^{N} (y_i - f(x_i))^2$$
$$= \frac{1}{2} * ((y_1 - f(x_1))^2 + (y_2 - f(x_2))^2)$$
$$= \frac{1}{2} * ((0.9 - f(2.5))^2 + (0.2 - f(0.5))^2)$$

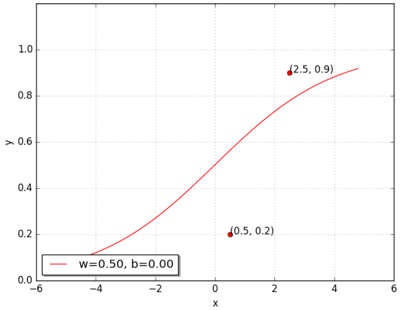


$$\mathscr{L}(w,b) = \frac{1}{2} * \sum_{i=1}^{N} (y_i - f(x_i))^2$$

$$= \frac{1}{2} * ((y_1 - f(x_1))^2 + (y_2 - f(x_2))^2)$$

$$= \frac{1}{2} * ((0.9 - f(2.5))^2 + (0.2 - f(0.5))^2)$$

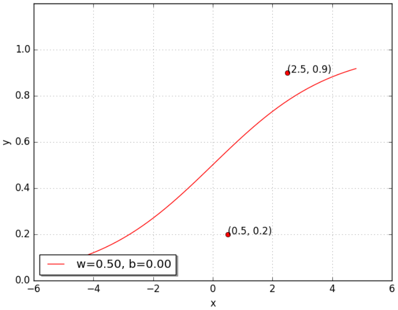
$$= 0.073$$



$$
\begin{aligned}
\mathscr{L}(w,b) &= \frac{1}{2} * \sum_{i=1}^{N} (y_i - f(x_i))^2 \\
&= \frac{1}{2} * ((y_1 - f(x_1))^2 + (y_2 - f(x_2))^2) \\
&= \frac{1}{2} * ((0.9 - f(2.5))^2 + (0.2 - f(0.5))^2) \\
&= 0.073
\end{aligned}
$$

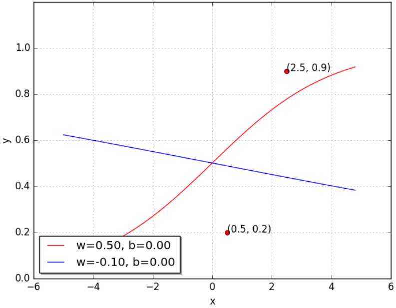
We want $\mathscr{L}(w,b)$ to be as close to 0 as possible

Let us try some other values of w, b



w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730

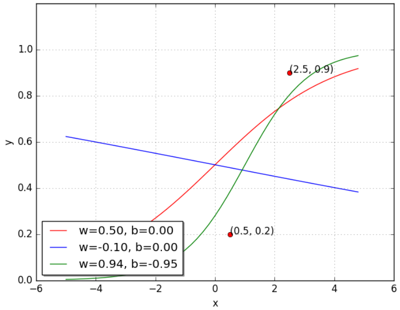
Let us try some other values of w, b



w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730
-0.10	0.00	0.1481

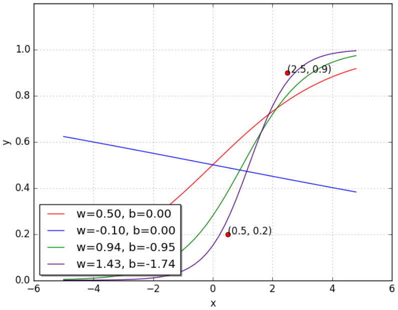
Oops!! this made things even worse...

Let us try some other values of w, b



w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730
-0.10	0.00	0.1481
0.94	-0.94	0.0214

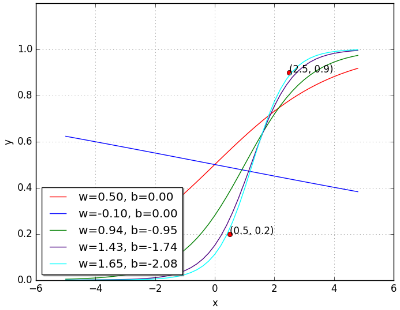
Perhaps it would help to push w and b in the other direction...



Let us try some other values of w, b

w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730
-0.10	0.00	0.1481
0.94	-0.94	0.0214
1.42	-1.73	0.0028

Let us keep going in this direction, *i.e.*, increase w and decrease b

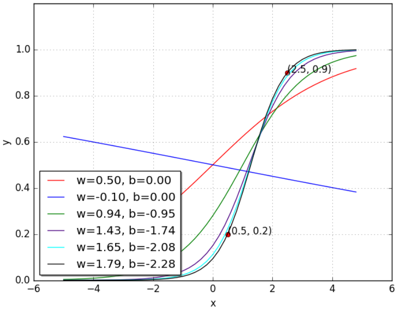


Let us try some other values of w, b

w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730
-0.10	0.00	0.1481
0.94	-0.94	0.0214
1.42	-1.73	0.0028
1.65	-2.08	0.0003

Let us keep going in this direction, *i.e.*, increase w and decrease b

Let us try some other values of w, b

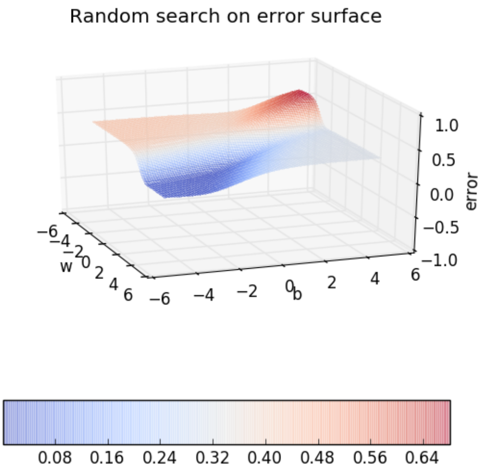


w	b	$\mathscr{L}(w,b)$
0.50	0.00	0.0730
-0.10	0.00	0.1481
0.94	-0.94	0.0214
1.42	-1.73	0.0028
1.65	-2.08	0.0003
1.78	-2.27	0.0000

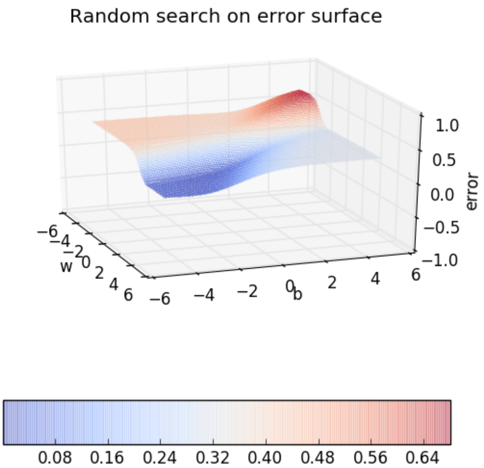
With some guess work and intuition we were able to find the right values for w and b

Let us look at something better than our “guess work” algorithm....

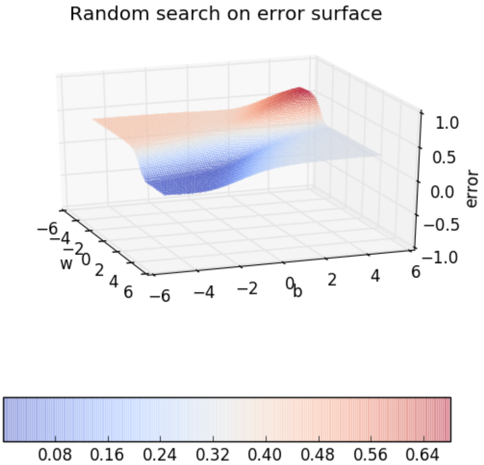
- Since we have only 2 points and 2 parameters $(w,\ b)$ we can easily plot $\mathscr{L}(w,b)$ for different values of $(w,\ b)$ and pick the one where $\mathscr{L}(w,b)$ is minimum



- Since we have only 2 points and 2 parameters $(w,\ b)$ we can easily plot $\mathscr{L}(w, b)$ for different values of $(w,\ b)$ and pick the one where $\mathscr{L}(w, b)$ is minimum

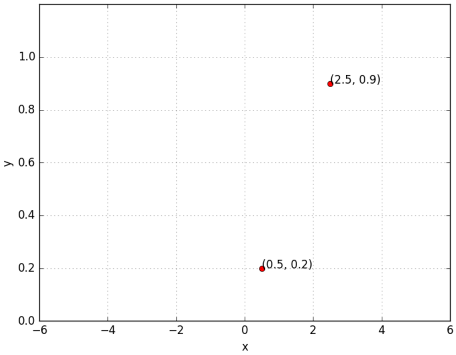


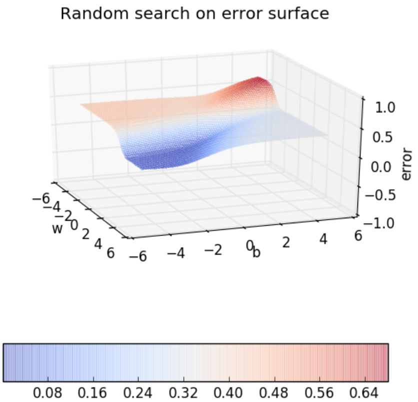
- Since we have only 2 points and 2 parameters $(w,\ b)$ we can easily plot $\mathscr{L}(w, b)$ for different values of $(w,\ b)$ and pick the one where $\mathscr{L}(w, b)$ is minimum
- But of course this becomes intractable once you have many more data points and many more parameters !!

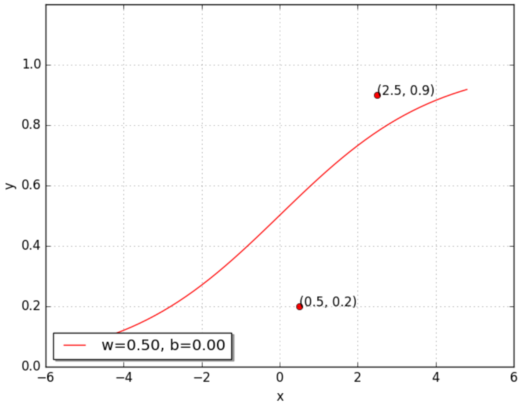


- Since we have only 2 points and 2 parameters (w, b) we can easily plot $\mathscr{L}(w, b)$ for different values of (w, b) and pick the one where $\mathscr{L}(w, b)$ is minimum
- But of course this becomes intractable once you have many more data points and many more parameters !!
- Further, even here we have plotted the error surface only for a small range of (w, b) [from $(-6, 6)$ and not from $(-\inf, \inf)$]

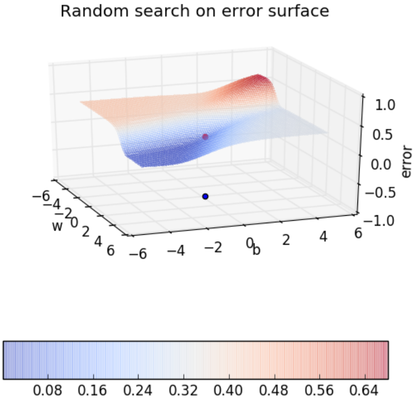
Let us look at the geometric interpretation of our "guess work" algorithm in terms of this error surface


(2.5, 0.9)
(0.5, 0.2)
y
x
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6

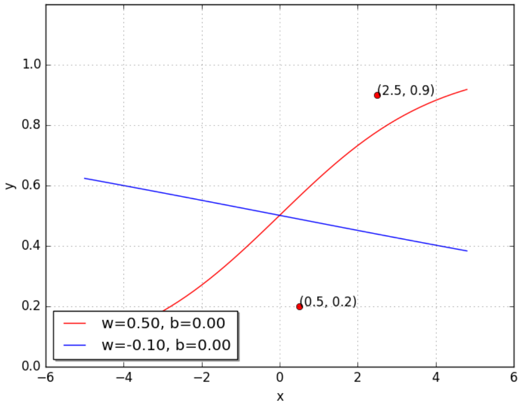

Random search on error surface
error
w
b
1.0
0.5
0.0
−0.5
−1.0
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64



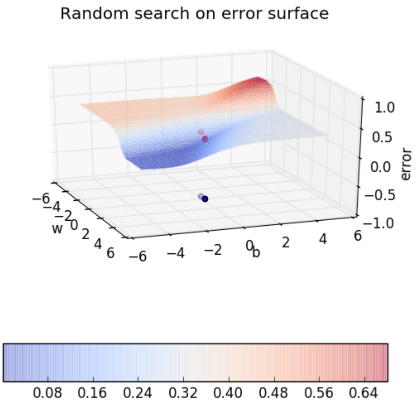
(2.5, 0.9)
(0.5, 0.2)
w=0.50, b=0.00
y
x



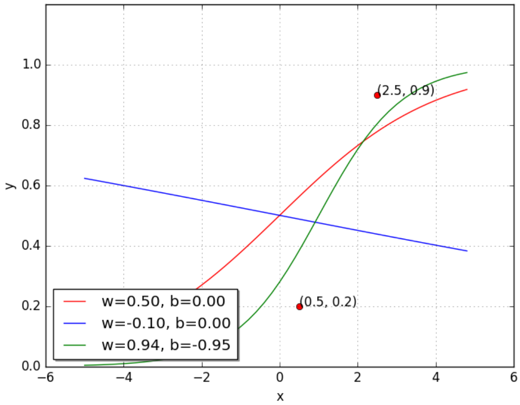
Random search on error surface
error
w
b
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64



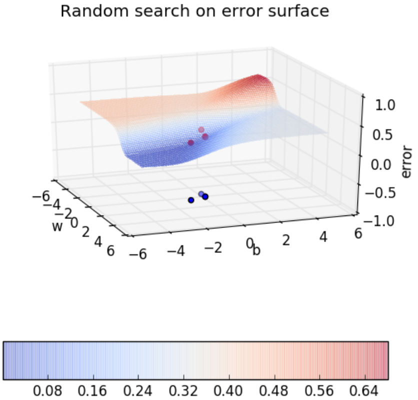
(2.5, 0.9)
(0.5, 0.2)
w=0.50, b=0.00
w=-0.10, b=0.00
y
x



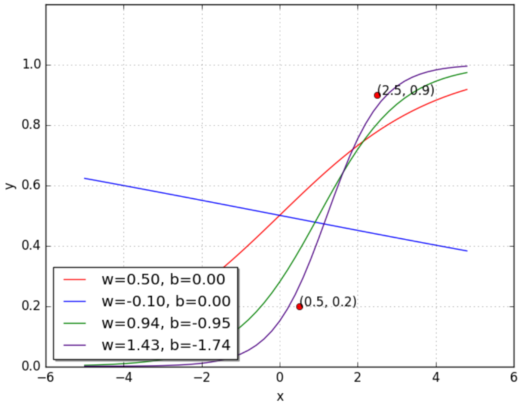
Random search on error surface
error
w
b
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64



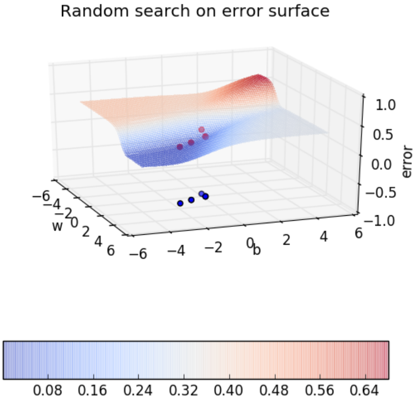
(2.5, 0.9)
(0.5, 0.2)
w=0.50, b=0.00
w=-0.10, b=0.00
w=0.94, b=-0.95
y
x
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6



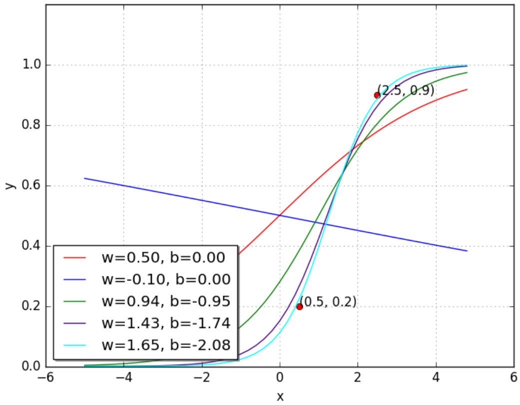
Random search on error surface
error
w
b
1.0
0.5
0.0
−0.5
−1.0
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64

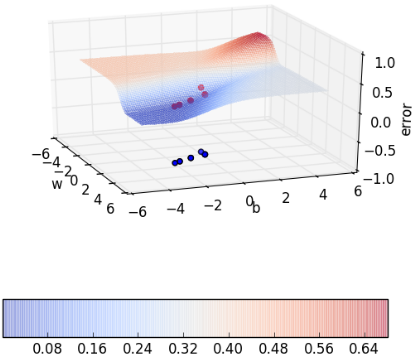


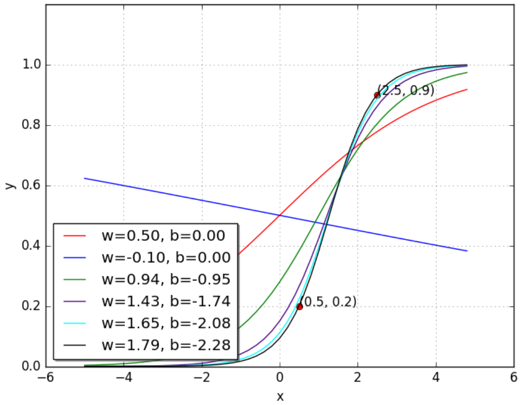
1.0
0.8
0.6
0.4
0.2
0.0
y
−6
−4
−2
0
2
4
6
x
(2.5, 0.9)
(0.5, 0.2)
w=0.50, b=0.00
w=-0.10, b=0.00
w=0.94, b=-0.95
w=1.43, b=-1.74



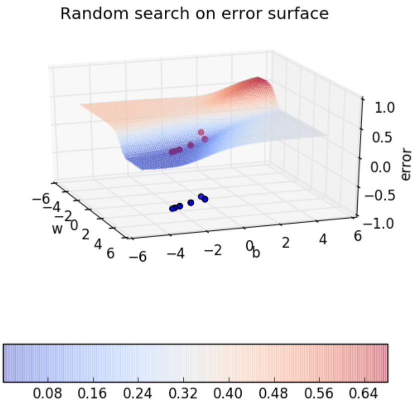
Random search on error surface
1.0
0.5
0.0
−0.5
−1.0
error
w
b
−6
−4
−2
0
2
4
6
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64


(2.5, 0.9)
(0.5, 0.2)
w=0.50, b=0.00
w=-0.10, b=0.00
w=0.94, b=-0.95
w=1.43, b=-1.74
w=1.65, b=-2.08
x
y


Random search on error surface
error
w
b
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64



w=0.50, b=0.00
w=-0.10, b=0.00
w=0.94, b=-0.95
w=1.43, b=-1.74
w=1.65, b=-2.08
w=1.79, b=-2.28
(2.5, 0.9)
(0.5, 0.2)
x
y



Random search on error surface
error
w
b
0.08
0.16
0.24
0.32
0.40
0.48
0.56
0.64