Convex Optimization

[1] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Available Online:
https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

Convex Sets

“In Euclidean space, a region is a convex set if the following is true. For any two points inside the region, a straight line segment can be drawn. If every point on that segment is inside the region, then the region is convex.”

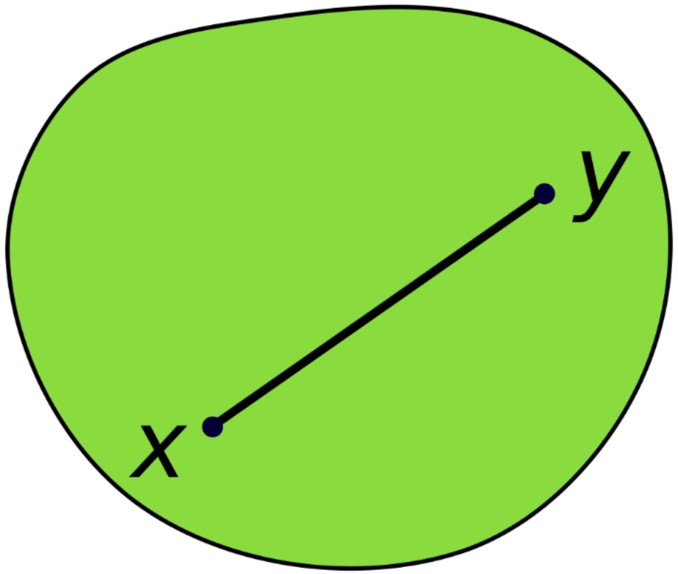
(Source: https://simple.wikipedia.org/wiki/Convex_set)

“A region is any connected part of a space or surface”

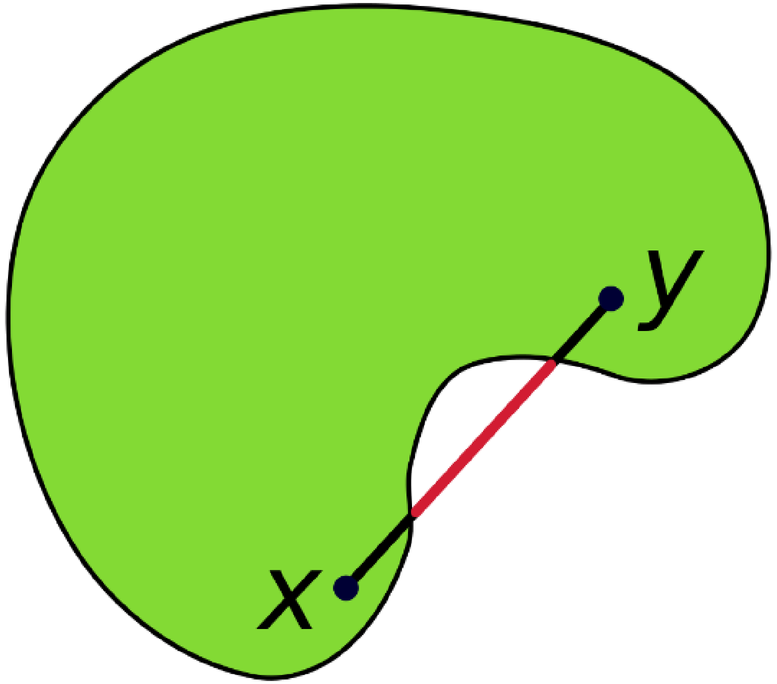
(Source; https://simple.wiktionary.org/wiki/region)

Convex Sets (cont'd)

Convex set



Non-Convex set

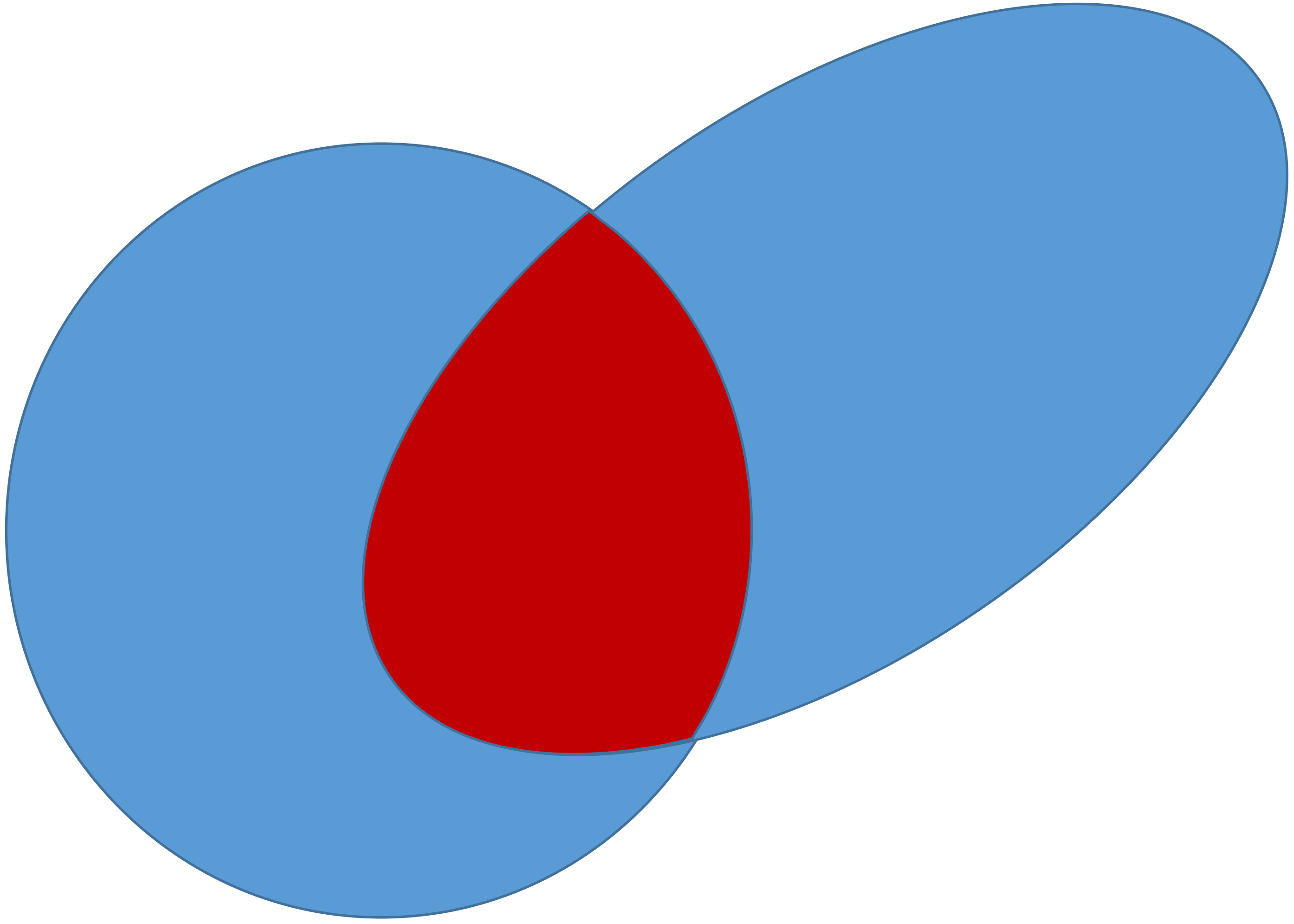


Images credit: https://simple.wikipedia.org/wiki/Convex_set

Intersection of two convex sets

Prove that intersection of convex sets is also a convex set.

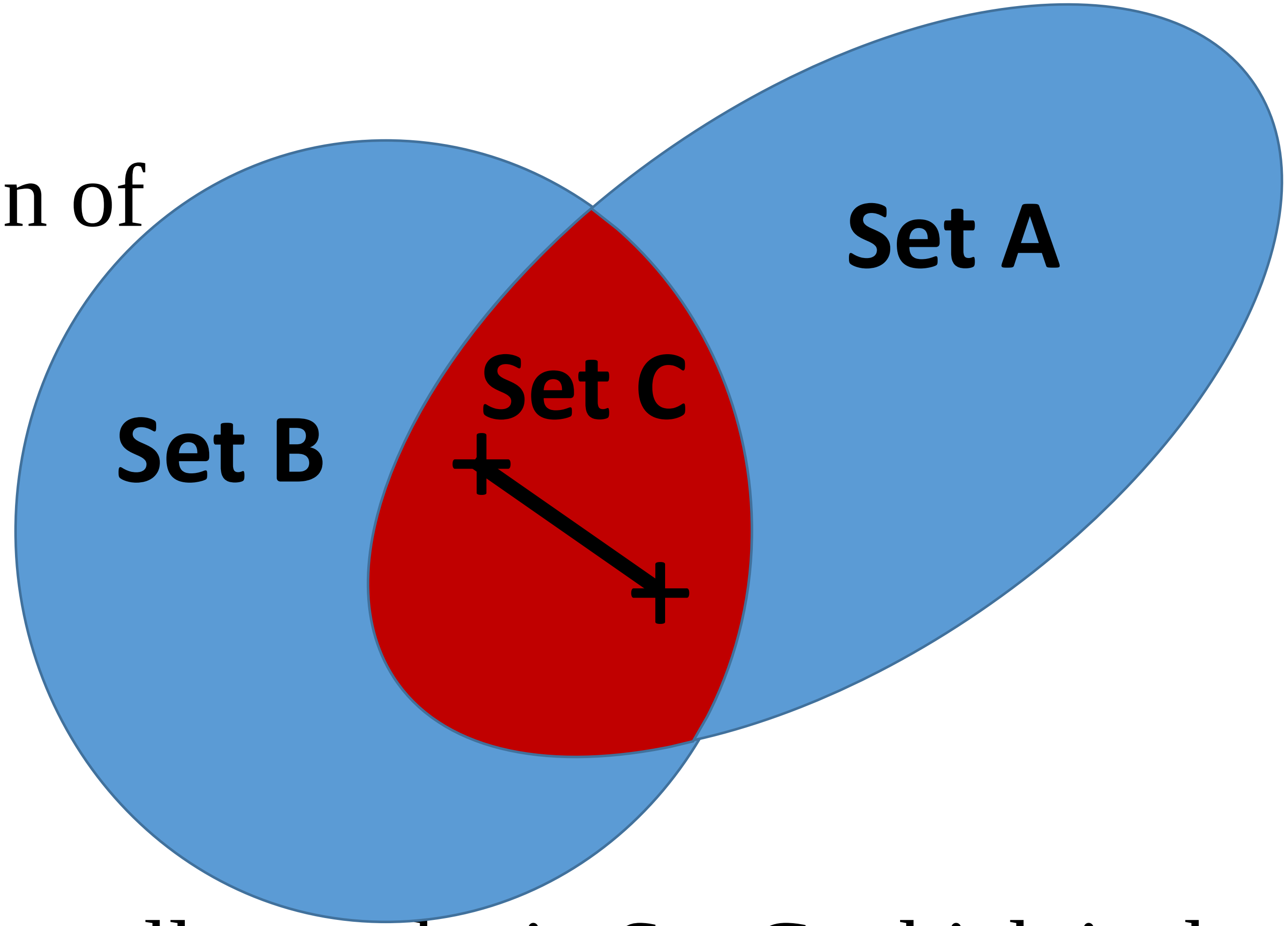
Intersection of two convex sets (cont'd)



Intersection of two convex sets (cont'd)

Proof:

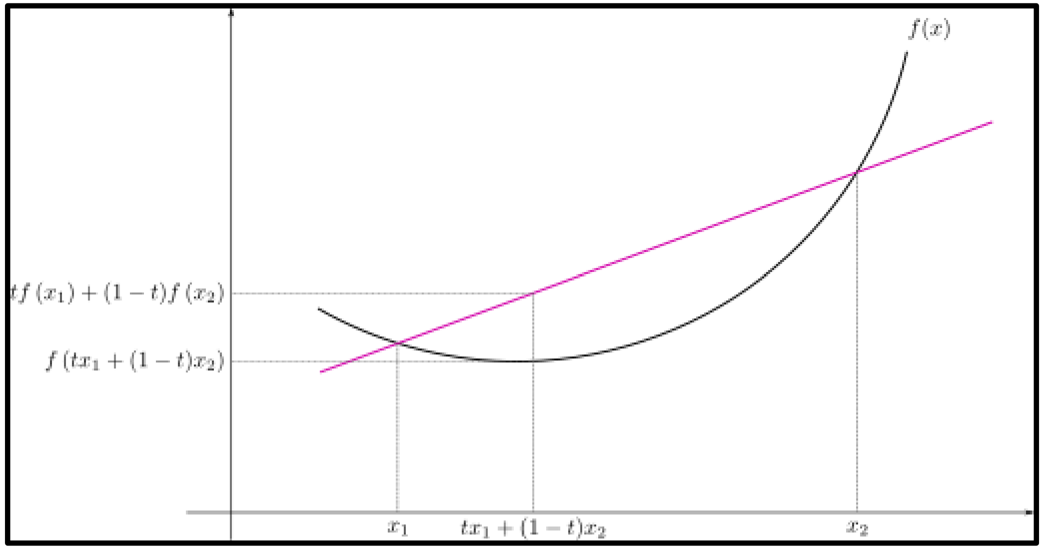
Let Set C be the intersection of Sets A and B. Consider any two points in Set C. These two points must be in Set A and also in Set B, so all the points between them must be in Set A and also in Set B. Therefore, they all must be in Set C which is the intersection of A and B.



QED

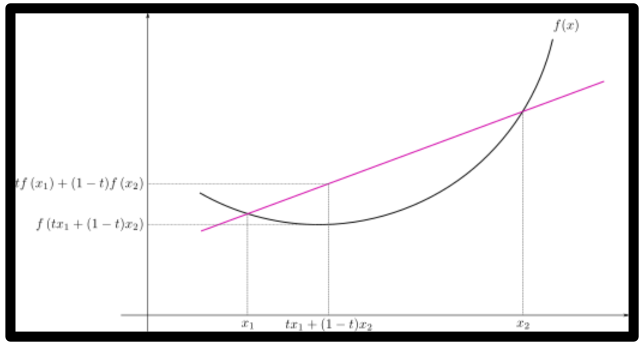
Convex Function

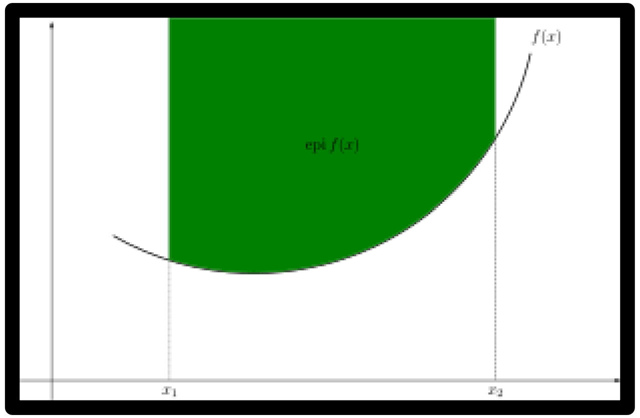
A real-valued function defined on $\mathbb{R}^n$ is called convex if the line segment between any two points on the graph of the function lies above or on the graph.



Source: https://en.wikipedia.org/wiki/Convex_function

Relationship between convex function and convex set





A function *f(x)* is convex if and only if the region above its graph, called the *epigraph of f(x)* or epi *f(x)*), is a convex set.

Source: https://en.wikipedia.org/wiki/Convex_function

Convex Function

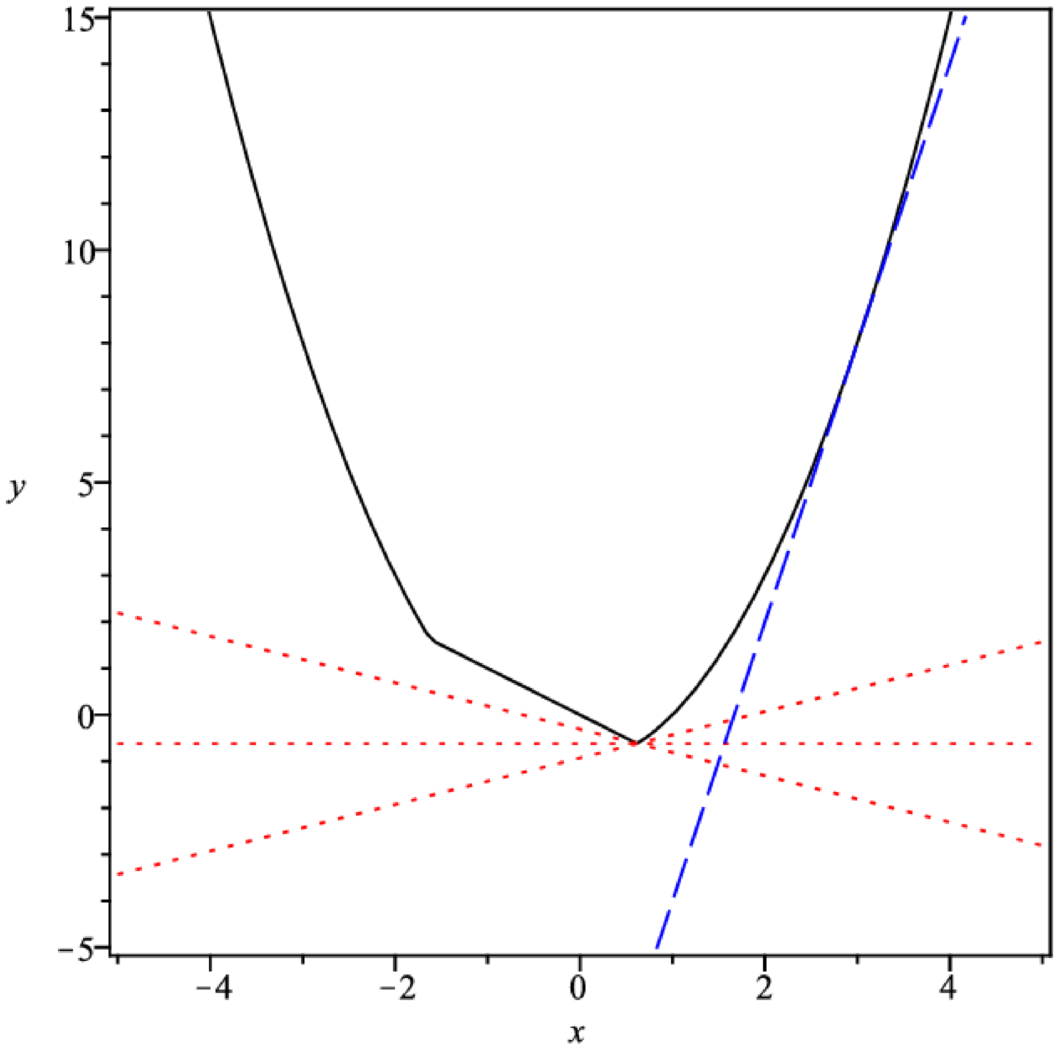


Image credit: Jennifer Johnstone, *Convexity of the Proximal Average,* BSc Thesis, The University of British Columbia, Canada, 2008.

.

Convex Function

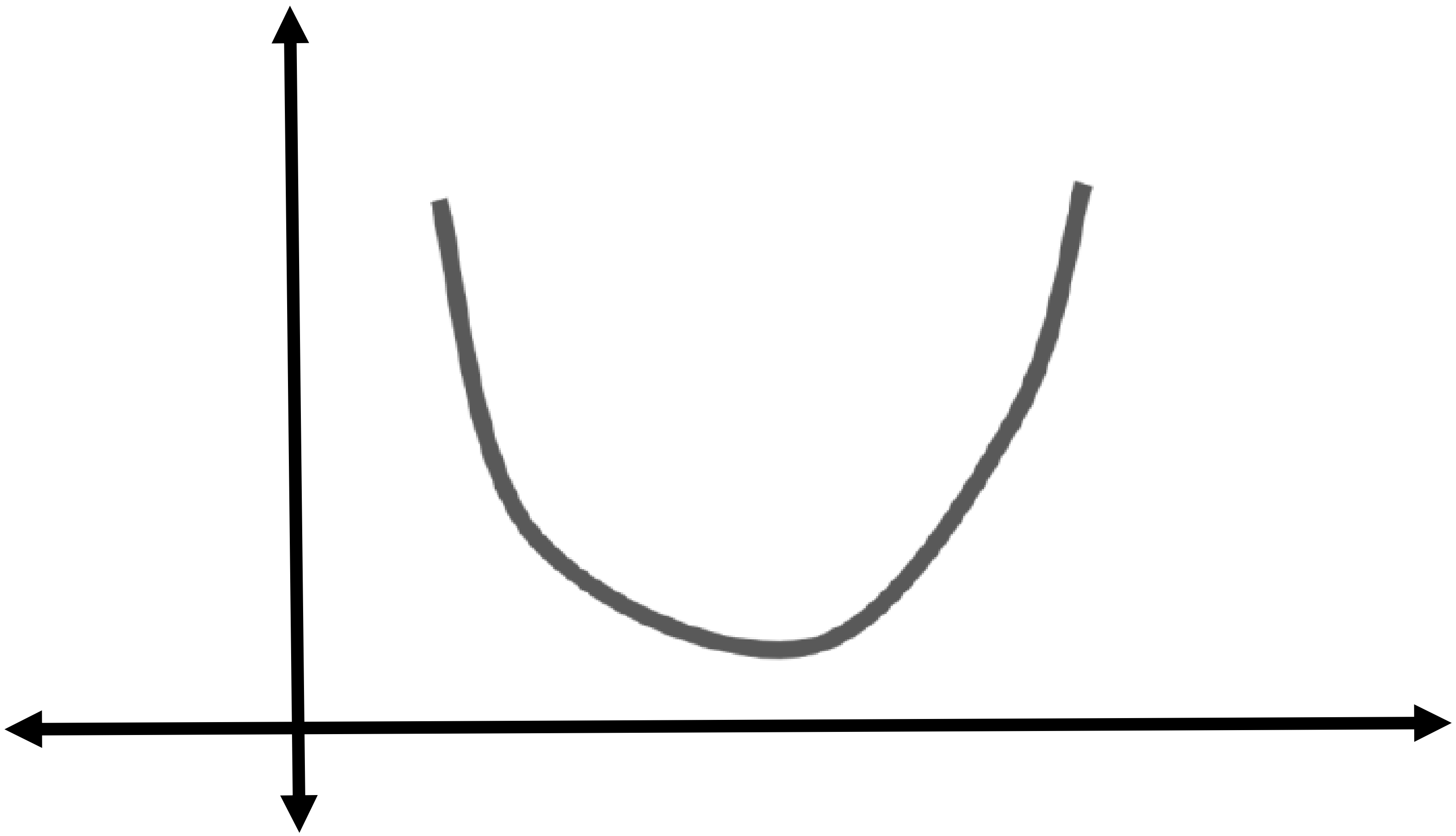
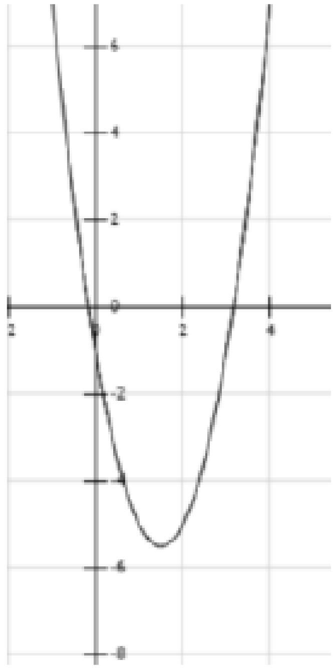


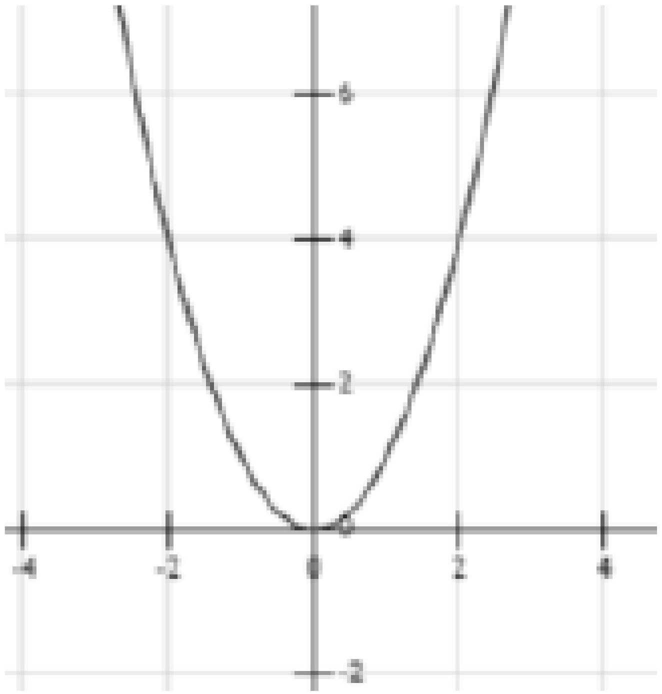
Image credit: https://developers.google.com/machine-learning/crash-course/reducing-loss/gradient-descent

More Examples of Convex Functions

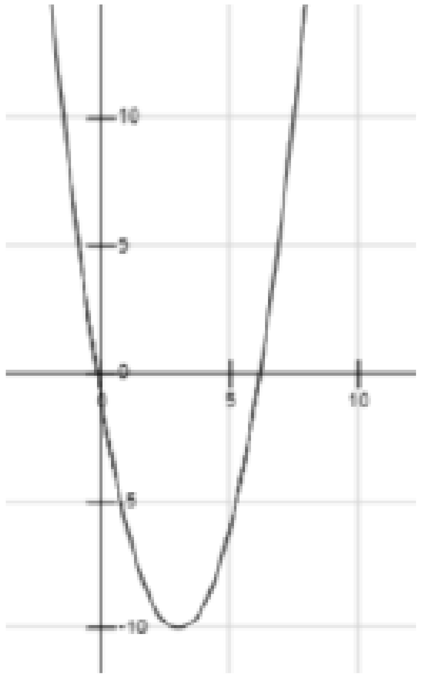
$f(x) = x^2 + (x-3)^2 - 10$



$f(x) = x^2$



$f(x) = (x-3)^2 - 10$



Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

Concave Function

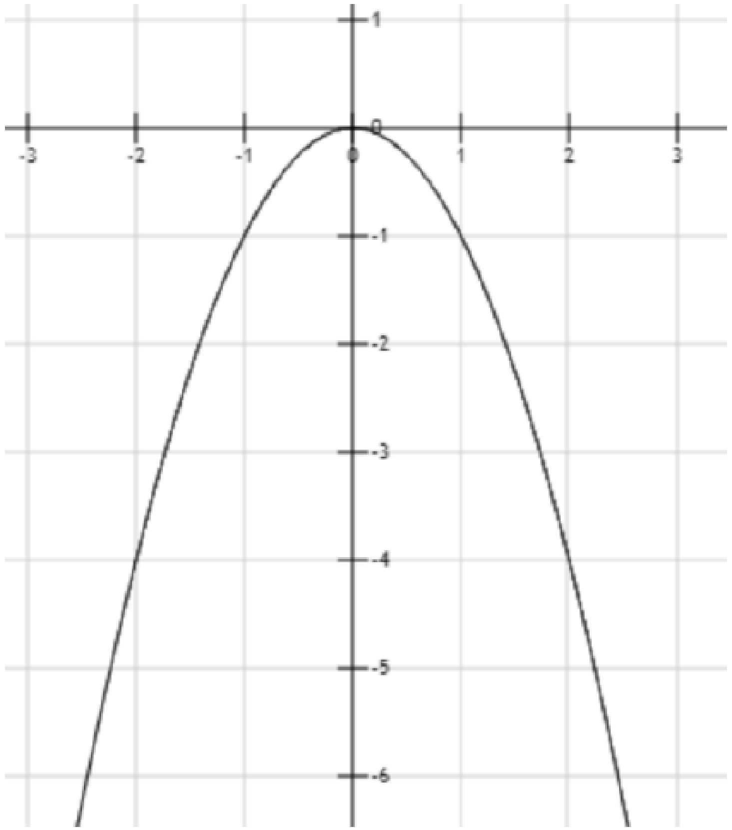
A function $f(x)$ is concave if there exists a function $g(x)$ which is convex and

$$f(x) = -g(x)$$

In other words, a concave function is the negative of a convex function.

Concave Function

Example: $f(x) = -x^2$



Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ-

Concave Function

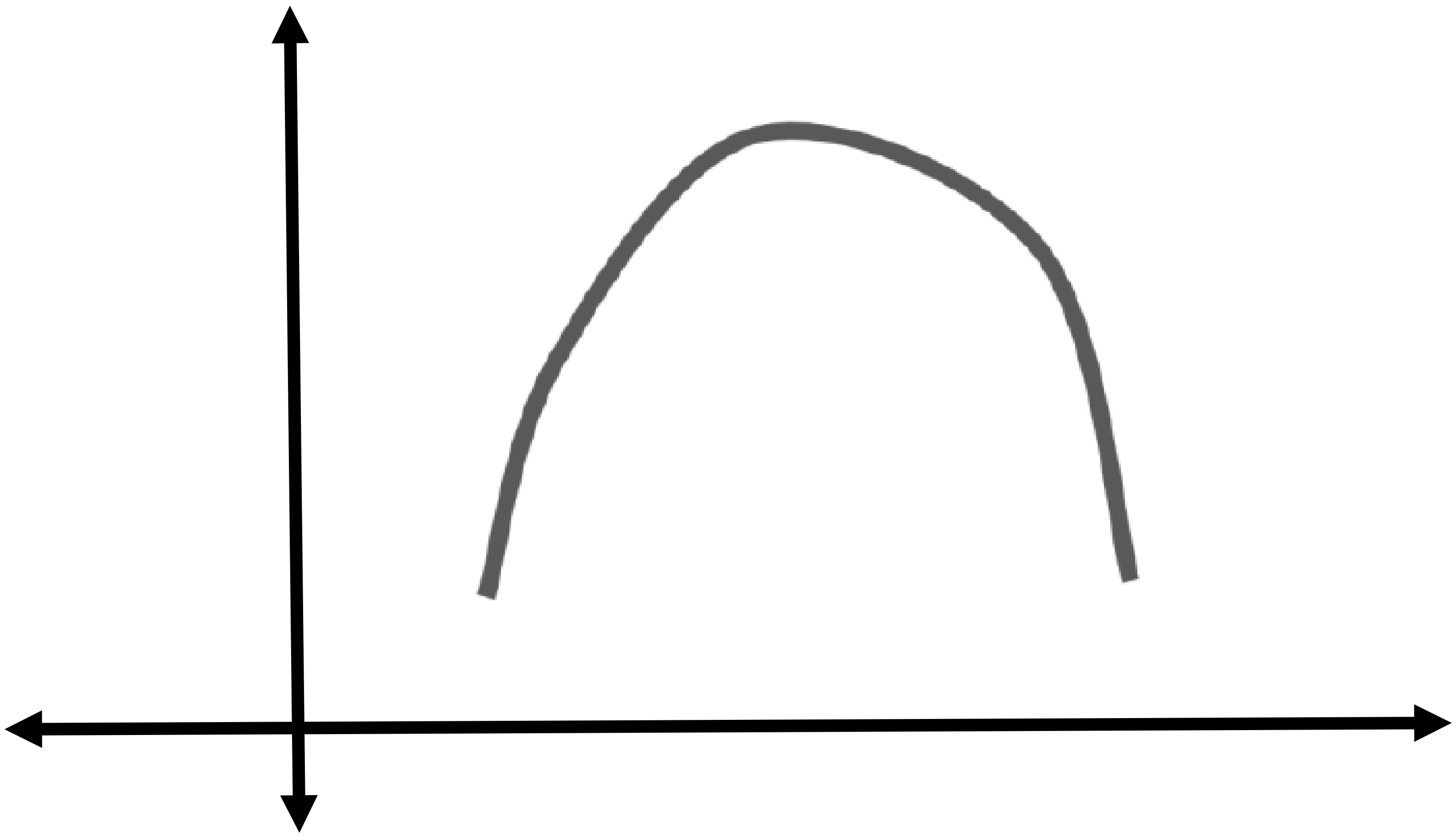
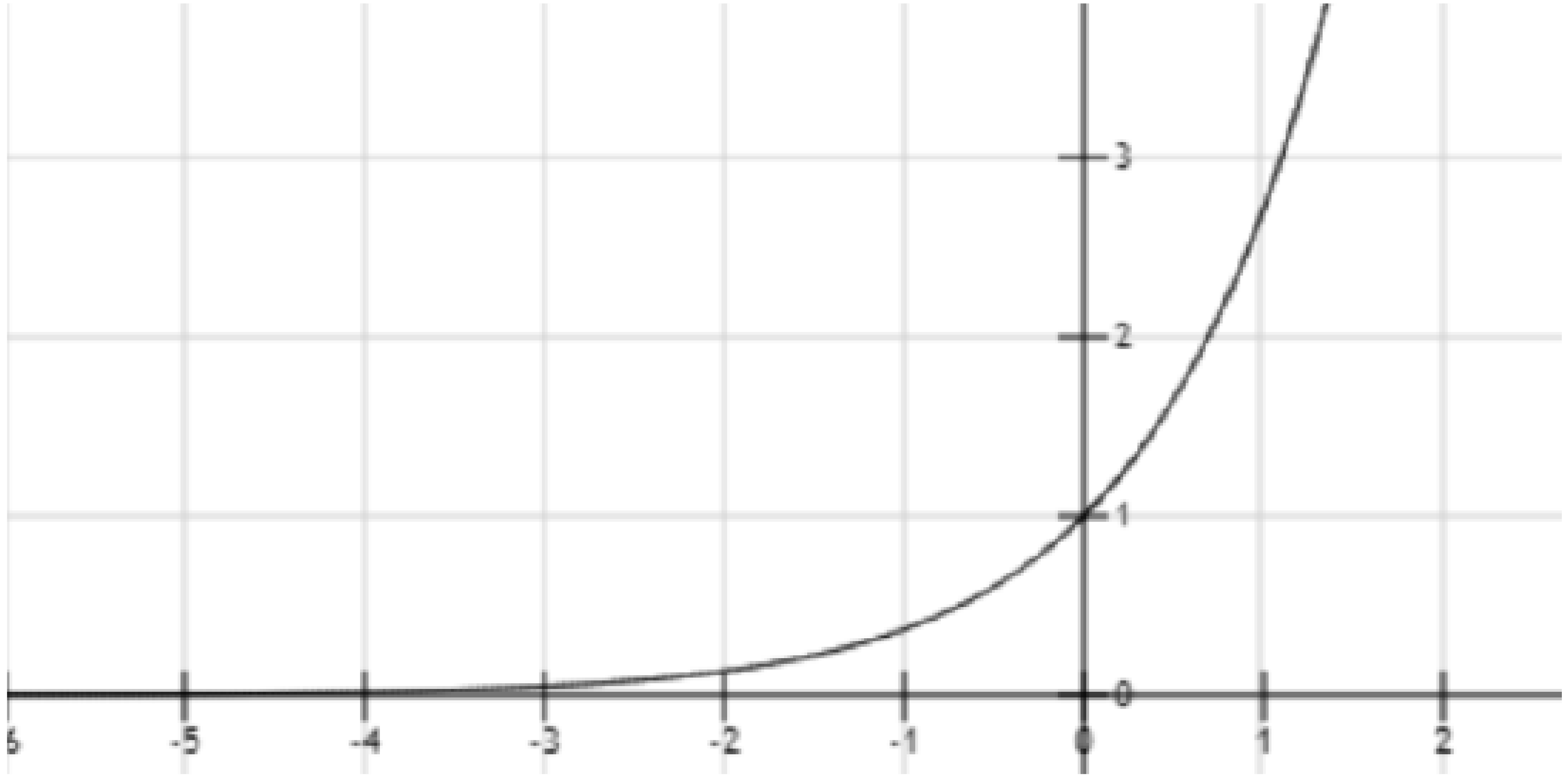


Image credit: https://developers.google.com/machine-learning/crash-course/reducing-loss/gradient-descent

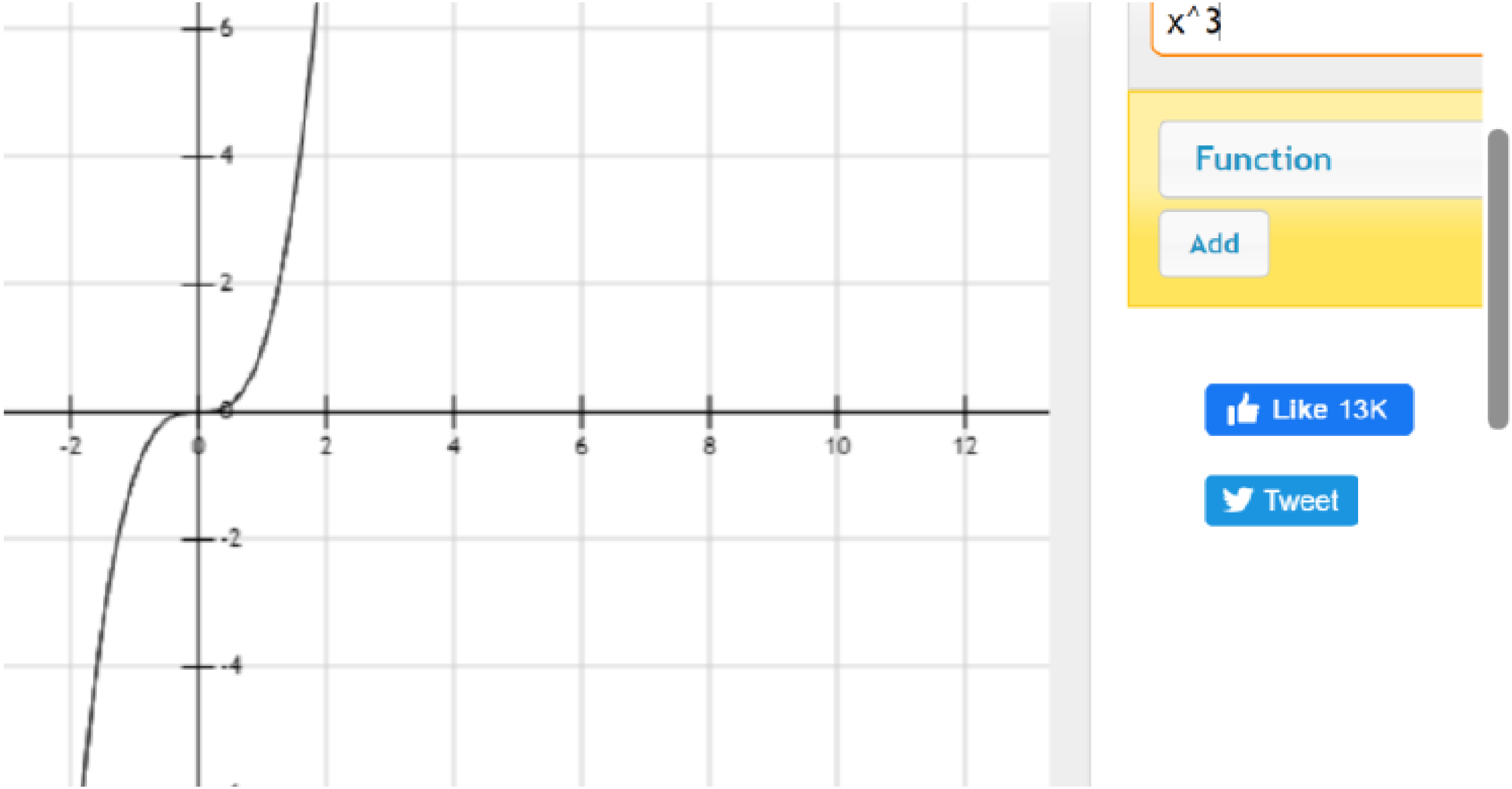
Another example of a convex function

$f(x) = e^x$ is a convex function



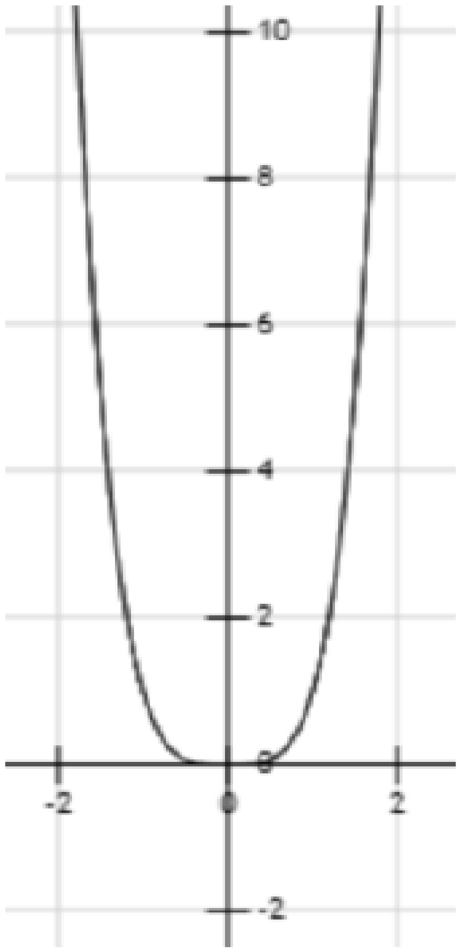
Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

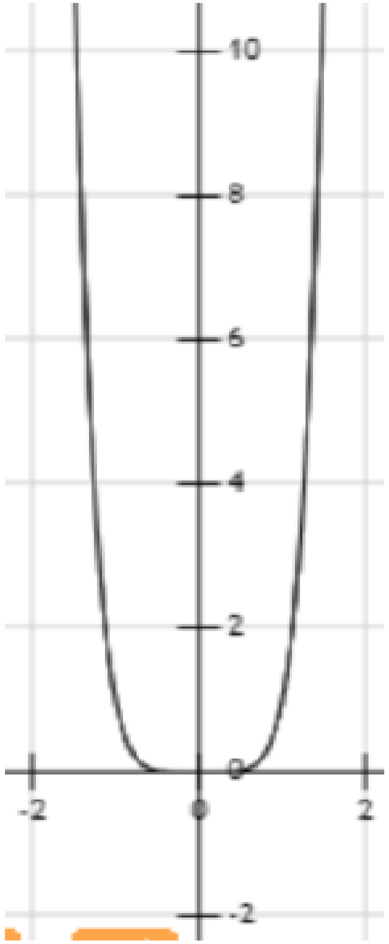
$f(x) = x^3$ is non-convex. It is convex for $x \geq 0$, and concave for $x \leq 0$.

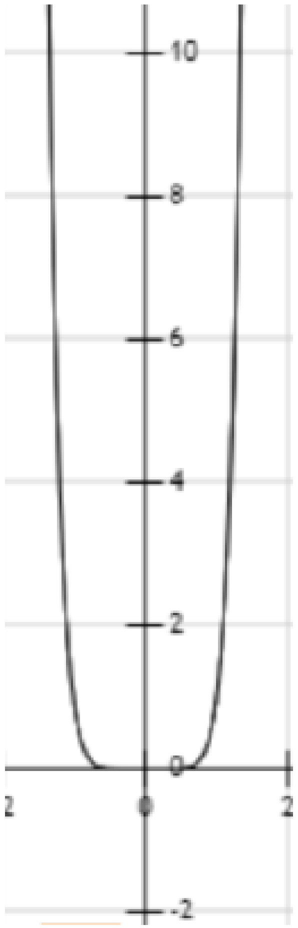


Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

$f(x) = x^4$ is convex $f(x) = x^6$ is convex $f(x) = x^8$ is convex







[Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--](http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--)

The functions

$f(x) = x^5$

and

$f(x) = x^7$

are non-convex.

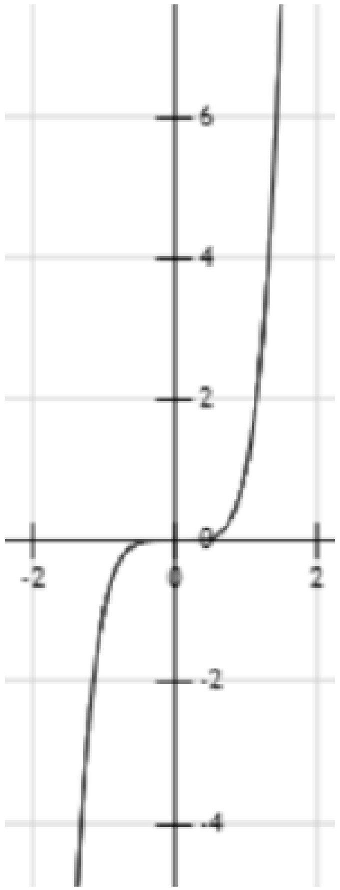
They are convex for

$x \geq 0,$

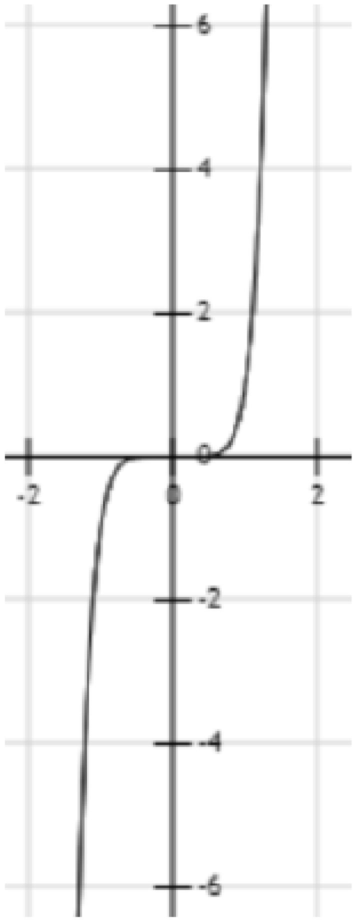
and concave for

$x \leq 0.$

$f(x) = x^5$



$f(x) = x^7$



Credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

$$f(x) = x^n$$

$f(x) = x^n$ is convex for $n = 1$ and for n even.

For $n > 1$ odd,
it is convex for $x \geq 0$
and concave for $x \leq 0$.

Let's prove that $y = x^2$ is convex.

To prove that $y = x^2$ is convex, we need to show that the points on the line segment between the points (x_1, y_1) and (x_2,y_2) are above or on the function $y = x^2$.

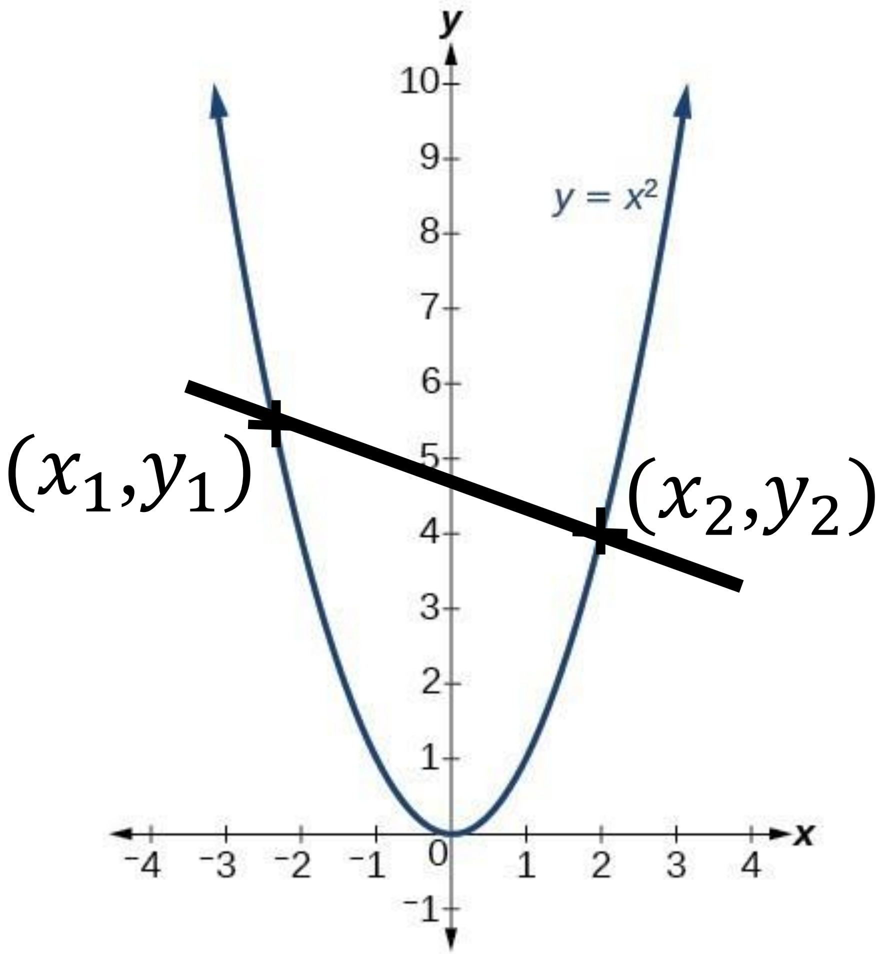


Image credit: https://courses.lumenlearning.com/waymakercollegealgebra/chapter/transformations-of-quadratic-functions/

Let's prove that $y = x^2$ is convex (cont'd)

In particular, to prove that $y = x^2$ is convex, we need to show that for any point (x_3, y_3) on the line segment between the points (x_1, y_1) and (x_2, y_2), $y_3 \geq (x_3)^2$.

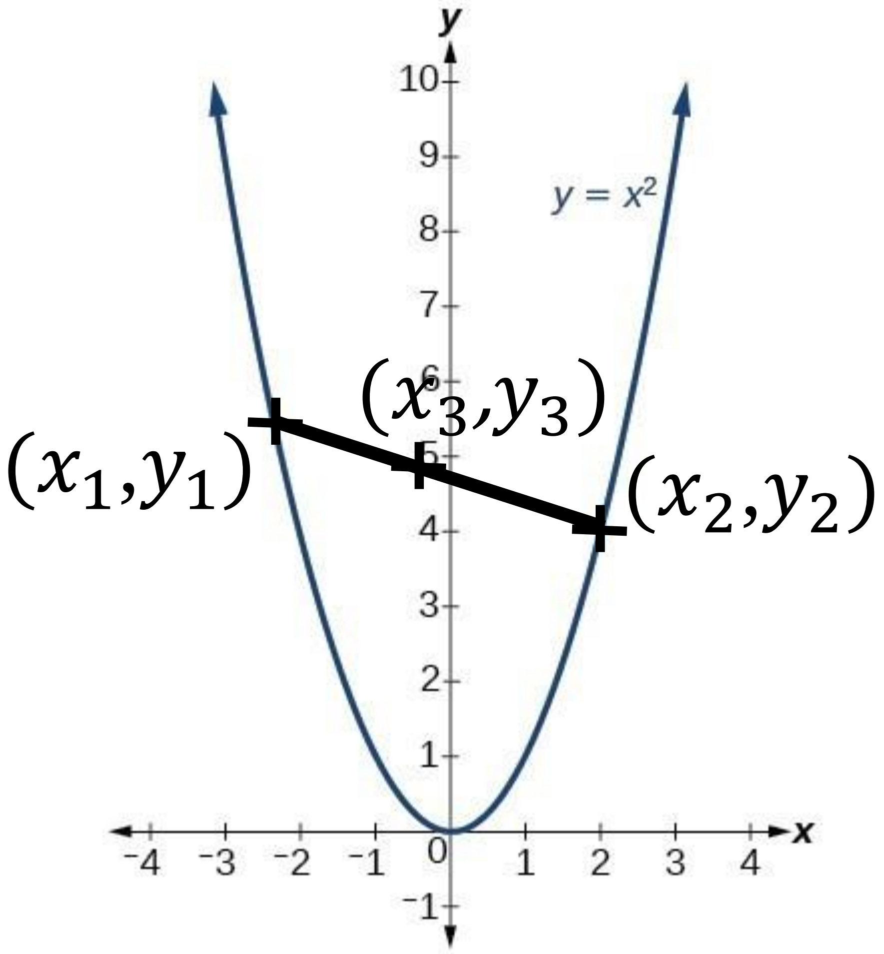


Image credit: https://courses.lumenlearning.com/waymakercollegealgebra/chapter/transformations-of-quadratic-functions/

Let's prove that $y = x^2$ is convex (cont'd)

$x_3 = px_1 + (1-p)x_2$

$y_3 = py_1 + (1-p)y_2$

For $0 \le p \le 1$.

$y_3 = p(x_1)^2 + (1-p)(x_2)^2$

$(x_3)^2 = [px_1 + (1-p)x_2]^2$.

We need to show:

$p(x_1)^2 + (1-p)(x_2)^2 \ge [px_1 + (1-p)x_2]^2$

or

$p(x_1)^2 + (1-p)(x_2)^2 \ge p^2(x_1)^2 + 2p(1-p)x_1x_2 + (1-p)^2(x_2)^2$

Let’s prove that $y = x^2$ is convex (cont’d)

or

$p(x_1)^2+(x_2)^2 - p(x_2)^2 \geq p^2(x_1)^2+2p(1-p)x_1x_2+(x_2)^2 - 2p(x_2)^2 + p^2(x_2)^2$

or

$p(x_1)^2 \geq p^2(x_1)^2+2p(1-p)x_1x_2 - p(x_2)^2 + p^2(x_2)^2$

or

$0 \geq (p^2-p)(x_1)^2+2p(1-p)x_1x_2 + (p^2-p)(x_2)^2$

or

$$0 \geq (p^2-p)(x_1)^2 - 2(p^2-p)x_1x_2 + (p^2-p)(x_2)^2$$

or

$0 \geq (p^2-p)(x_1 - x_2)^2$.

which must be true as $(p^2-p) \leq 0$ and $(x_1 - x_2)^2 \geq 0$.   **QED.**

A linear function

$$f(x) = ax + b$$

Where a and b are scalars is **both convex and concave.** Clearly for any two points on the line $f(x) = ax + b$, all the points in the segment between these two points are on the line, so $f(x)$ is convex. Then, -$f(x)$ is also convex because it is also a linear function that satisfies The above form so $f(x)$ is also concave.

Operations that Preserve Convexity

The operations that we discuss apply to $x \in \mathbb{R}^n$,
but most of our examples are for $x \in \mathbb{R}$.

It is given that $f_1(x)$ and $f_2(x)$ are convex functions.

Prove that the function
$f(x) = f_1(x) + f_2(x)$
is also convex.

Proof

We need to show:

$f(px_1$+(1-p)$x_2)\leq pf(x_1)$+(1-p) $f(x_2)$

We know that f_1 and f_2 are convex so

$f_1(px_1$+(1-p)$x_2)\leq pf_1(x_1)$+(1-p)$f_1(x_2)$
$f_2(px_1$+(1-p)$x_2)\leq pf_2(x_1)$+(1-p)$f_2(x_2)$
and
$f(x) = f_1(x) + f_2(x)$.

Proof (cont'd)

$f(px_1+(1\text{-p})x_2)$

$= f_1(px_1+(1-\text{p})x_2) + f_2(px_1+(1-\text{p})x_2)$ [because $f = f_1+f_2$]

$\leq pf_1(x_1)+(1-\text{p})f_1(x_2)+pf_2(x_1)+(1\text{-p})f_2(x_2)$ [because f_1 and f_2 are convex]

$= p(f_1(x_1)+f_2(x_1))+(1-\text{p})(f_1(x_2)+f_2(x_2))$

$= pf(x_1)+(1\text{-p})\,f(x_2)$. [because $f= f_1+f_2$.]

QED

Operations that Preserve Convexity (Cont'd)

Nonnegative Weighted Sum

If $f_1(x), f_2(x), f_3(x), \ldots, f_m(x)$ are convex functions and $w_1, w_2, w_3, \ldots, w_m$ are nonnegative scalars, then $f(x) = \sum_{i=1}^{m} w_i f_i(x)$ is a convex function.
This is an extension of the convexity of sum of convex functions discussed earlier. Prove it as homework.

Credit: http://www.princeton.edu/~aaa/Public/Teaching/ORF363_COS323/F14/ORF363_COS323_F14_Lec6.pdf

Using the previous rule, prove that a nonnegative weighted sum of concave functions is concave. That is, prove that if $f_1(x), f_2(x), f_3(x), \ldots, f_m(x)$ are concave functions and $w_1, w_2, w_3, \ldots, w_m$ are nonnegative scalars, then
$f(x) = \sum_{i=1}^{m} w_i f_i(x)$ is a concave function.

Proof

Let $g(x) = -f(x)$ and let
$g_i(x) = -f_i(x)$, i=1,2,3, ..., m

We know that $g_i(x)$, i=1,2,3, ..., m are convex by definition. Then, according to the first rule $g(x)$ is also convex. Therefore, by definition $f(x)$ is concave.

QED

Operations that Preserve Convexity (Cont'd)

Composition with linear mapping

If $f(x)$ is a convex function, then $f(ax + b)$ is a convex function, where a and b are scalars and assuming that the domain of $f(ax + b)$ is $\{x|ax + b \in \text{domain of } f(x)\}$.

Can you prove it?

(Note: this rule is extendable, when $f: \mathbb{R}^n \to \mathbb{R}$, to the symbol a being a matrix and b a vector.)

Credit: http://www.princeton.edu/~aaa/Public/Teaching/ORF363_COS323/F14/ORF363_COS323_F14_Lec6.pdf

Operations that Preserve Convexity (Cont'd)

Proof

Consider two points x_1 and x_2 that satisfy the domain requirements. Then, for $0 \leq p \leq 1$,

$f(a(px_1 + (1-p)x_2) + b)$

$= f(p(ax_1) + (1-p)ax_2) + pb + (1-p)b)$

$= f(p(ax_1 + b) + (1-p)(ax_2 + \mathrm{b}))$

$\leq pf\ (ax_1 + b) + (1-p)\ f\ (ax_2 + b)$ [recall the f is convex.]

QED

Credit: http://www.princeton.edu/~aaa/Public/Teaching/ORF363_COS323/F14/ORF363_COS323_F14_Lec6.pdf

Question

How about the function $f(x) = e^{-x}$

Is it a convex function? (Yes/No)

Clearly justify your answer.

Question

How about the function $f(x) = e^{5-6x} + 6x^6 - 5x$

Is it a convex function? (Yes/No)

Clearly justify your answer.

Question

Consider the function $f(x) = e^{5-6x} + 6x^6 + 5x$

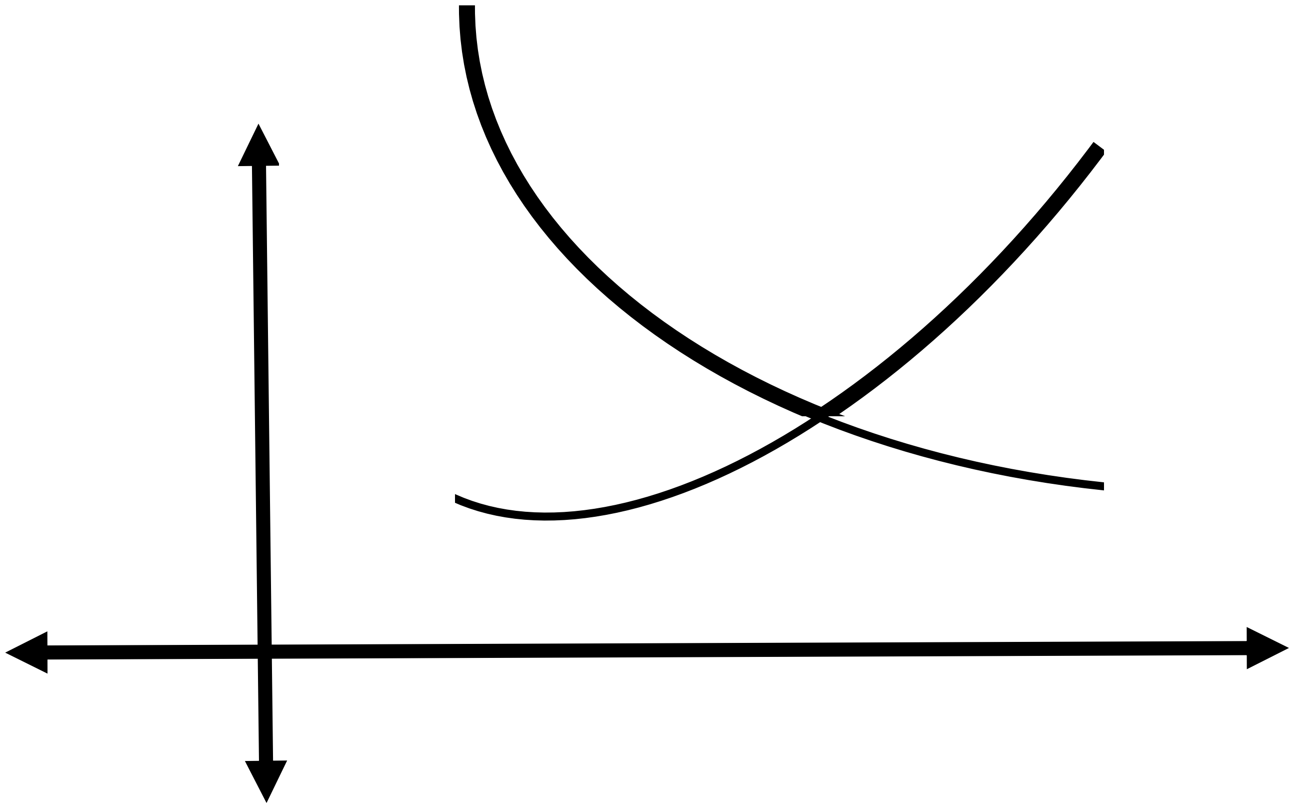
Is it a convex function? (Yes/No)

Clearly justify your answer.

Operations that Preserve Convexity (Cont'd)

Pointwise Maximum

$$f(x) = \max\{f_1(x), f_2(x)\}$$



Operations that Preserve Convexity (Cont'd)

Pointwise Maximum (cont'd)

Given that functions f_1, f_2, f_3, …, f_m are convex,
Then the function

$$f(x) = \max\{f_1(x), f_2(x), f_3(x), ..., f_m(x)\}$$

with a domain that is the intersection of the domains of f_1, f_2, f_3, …, f_m is convex.

Credit: http://www.princeton.edu/~aaa/Public/Teaching/ORF363_COS323/F14/ORF363 _COS323_F14_Lec6.pdf

Proof

Let x_1 and x_2 be two points in the domain of $f(x)$. Then

$f(px_1 + (1-p)x_2) = f_j(px_1 + (1-p)x_2)$ [This must be equal for some j so that f_j is the maximum of $\{f_1, \ldots, f_m\}$ at the point $px_1 + (1-p)x_2$.]
$\leq pf_j(x_1) + (1-p)f_j(x_2)$ [because f_j is convex]
$\leq p[\max(f_1(x_1), \ldots, f_m(x_1))] + (1-p)[\max(f_1(x_2), \ldots, f_m(x_2))]$
$= pf(x_1) + (1-p)f(x_2)$.

QED

What is Convex Optimization?

$f(\boldsymbol{x})$ is a convex objective function

$\boldsymbol{x} \in \mathbb{R}^n$**, C** is a convex set (a subset of $\mathbb{R}^n$).

Minimize $f(\boldsymbol{x})$

such that

$\boldsymbol{x} \in$ **C**

What is Convex Optimization? (another answer)

$f(\boldsymbol{x})$ is a concave objective function
$\boldsymbol{x} \in \mathbb{R}^n$, **C** is a convex set (a subset of $\mathbb{R}^n$).

Maximize $f(\boldsymbol{x})$
such that
$\boldsymbol{x} \in$ **C**

Some convex programming problems can be solved analytically, but many practical ones can't be solved analytically. Here are some simple homework problems.

Minimize $\quad f(x) = x^2 - 20x + 100$

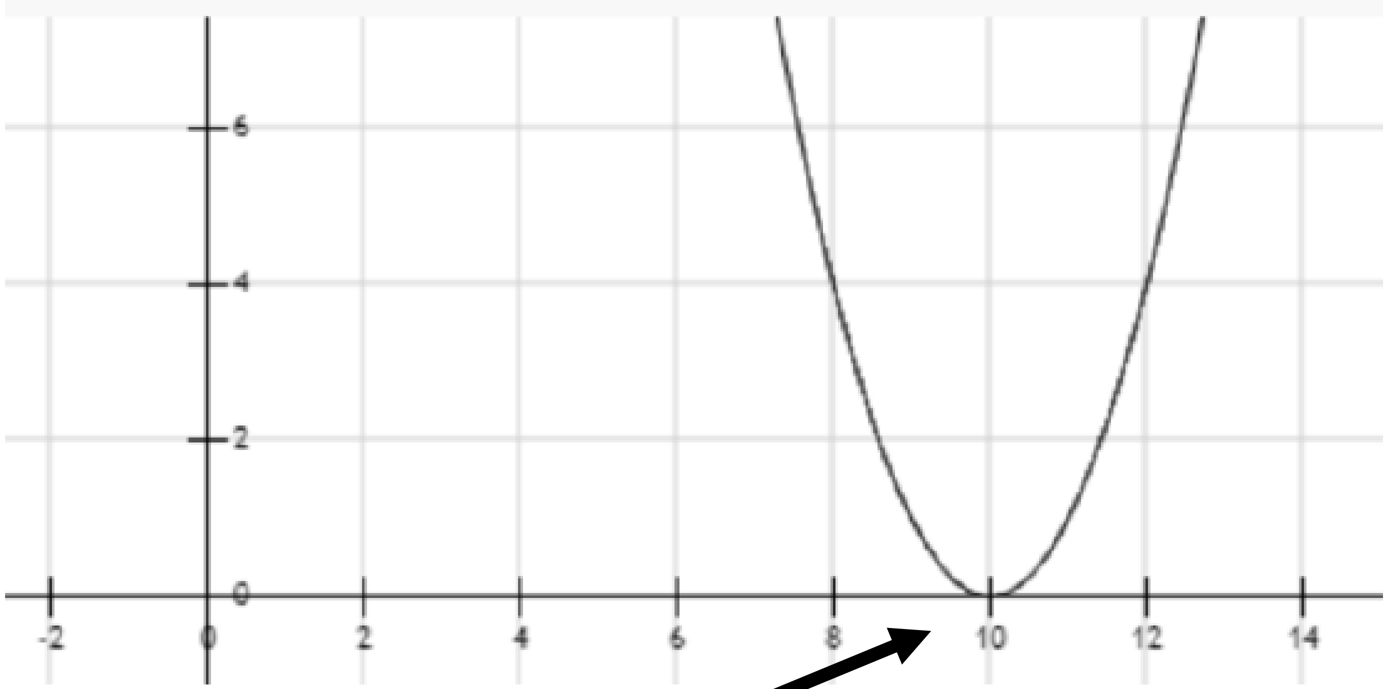
such that

$$2x \leq 50$$
$$x \geq 0.$$

Solve it analytically and by Excel.

(See Excel solution in the Excel file: convex_optimization Sheet1)

$$f(x) = x^2 - 20x + 100$$



Optimal solution $f(10) = 0$

Image credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

Minimize $f(x) = x^2 - 20x + 100$

such that

$$2x \leq 50$$

$$x \geq 12.$$

Solve it analytically and by Excel.

(See Excel solution in the Excel file: convex_optimization Sheet2)

Maximize $f(x) = 1 - 2^{-x}$

such that

$$x \leq 2$$

$$x \geq 0.$$

Solve it analytically and by Excel.

(See Excel solution in the Excel file: convex_optimization Sheet3)

Maximize $f(\boldsymbol{x}) = 2 - 2^{-x_1} - 50^{-x_2}$

such that

$$x_1 + x_2 \leq 2$$
$$\boldsymbol{x} \geq 0.$$

Solve it by Excel.

(See Excel solution in the Excel file: convex_optimization Sheet4)

Change the constraint to $x_1 + x_2 \leq 0.1$ and observe the changes in resource apportionment!!!

$$f(x) = 1 - 2^{-x}$$

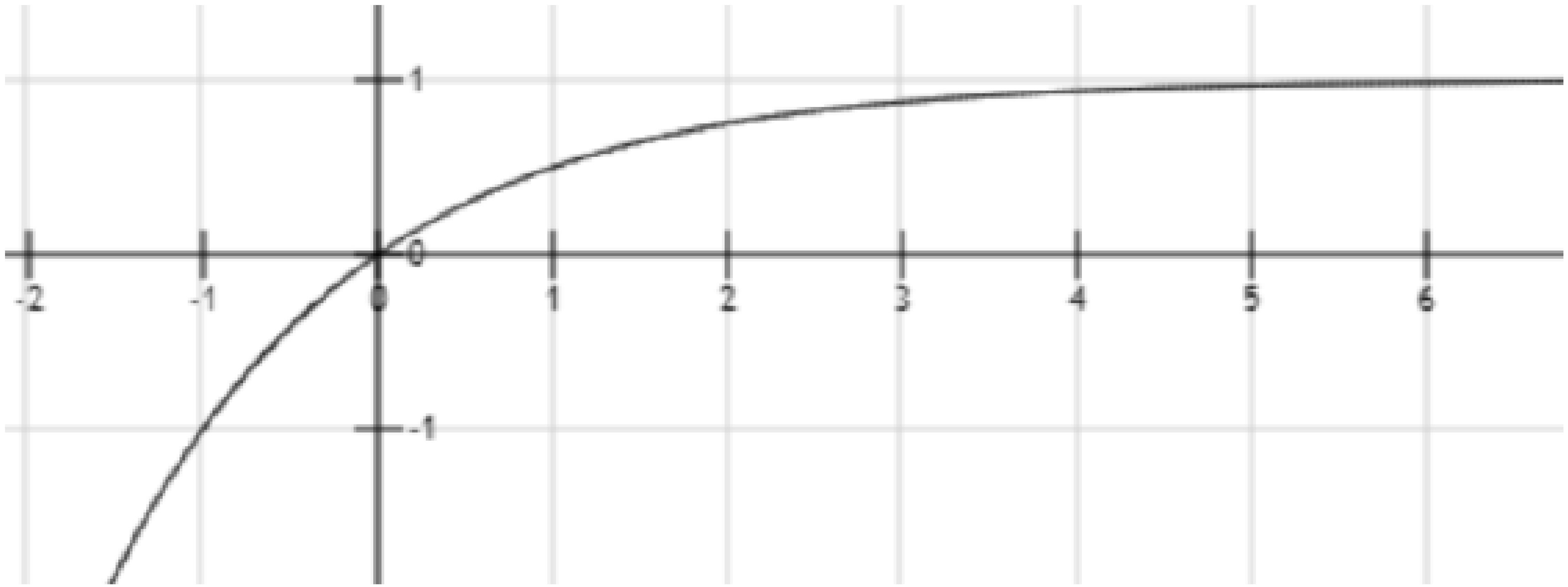


Image credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

$$f(x) = 1 - 50^{-x}$$

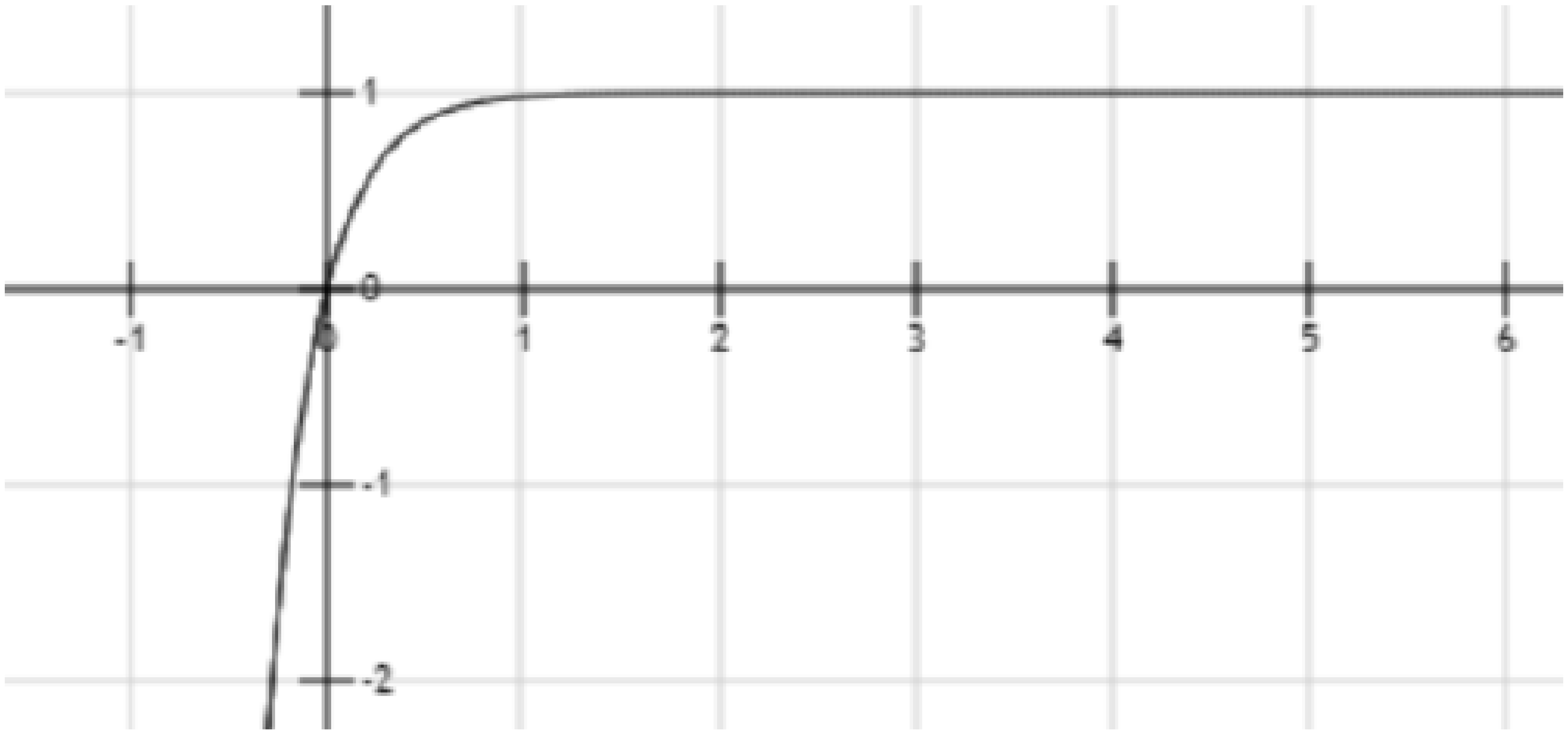


Image credit: http://fooplot.com/#W3sidHlwZSI6MCwiZXEiOiJ4XjIiLCJjb2xvciI6IiMwMDAwMDAifSx7InR5cGUiOjEwMDB9XQ--

Mathematical Optimization

A mathematical optimization problem has the form

Minimize $f_0(x)$

Subject to: $f_i(x) \leq b_i \quad i = 1, 2, 3, \dots, m.$

$x = (x_1, x_2, x_3, \ \dots, x_n)$ is a vector of decision variables, $x \in \mathbb{R}^n$.

The function $f_0 \ : \ \mathbb{R}^n \ \rightarrow \ \mathbb{R}$, is the objective function.

The functions $f_i \ : \ \mathbb{R}^n \ \rightarrow \ \mathbb{R}$ are the (inequality) constraint functions

The constants $b_1, b_2, b_3, \ \dots, b_n$ are the bounds of the constraint functions f_i $i = 1, 2, 3, \dots, m,$ respectively.

A vector $x^* \in \mathbb{R}^n$ is called optimal, or a solution to the above optimization problem, if it minimizes f_0(x) and satisfies the constraints.

That is, $f_i(x^*) \leq b_i$, and for any $z \in \mathbb{R}^n$ such that f_i(z)$\leq b_i, \ i = 1, 2, 3 \ \dots, m,$ we have that $f_0(x^*) \leq f_0$(z).

Source: [1].

In a **linear programming problem**,
the objective function f_0 and the constraint functions f_i, $i = 1, 2, 3, \ldots, m$, are linear. That is, f_i, $i = 0,1, 2, \ldots, m,$ satisfy

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y),$$

for any $x \in \mathbb{R}^n$, y $\in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$.

In a **convex programming (CP) problem,** they satisfy

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y).$$

Question: if a problem is LP, is it CP?

Source: [1]

Why do we focus on Convex Programming?

1. Convex programming has many applications including finance, production, economics, physics, biology, chemistry, internet resource allocation, optimal advertising*, data model fitting*, and political science. Note that there are more than 600,000 google scholar results to “convex optimization” (with the quotes).
2. It is much easier to solve than Non-Convex optimization.

* Credit: https://web.stanford.edu/~boyd/papers/pdf/cvx_applications.pdf

Non-Convex Optimization

Difficult to solve because of multiple local minima

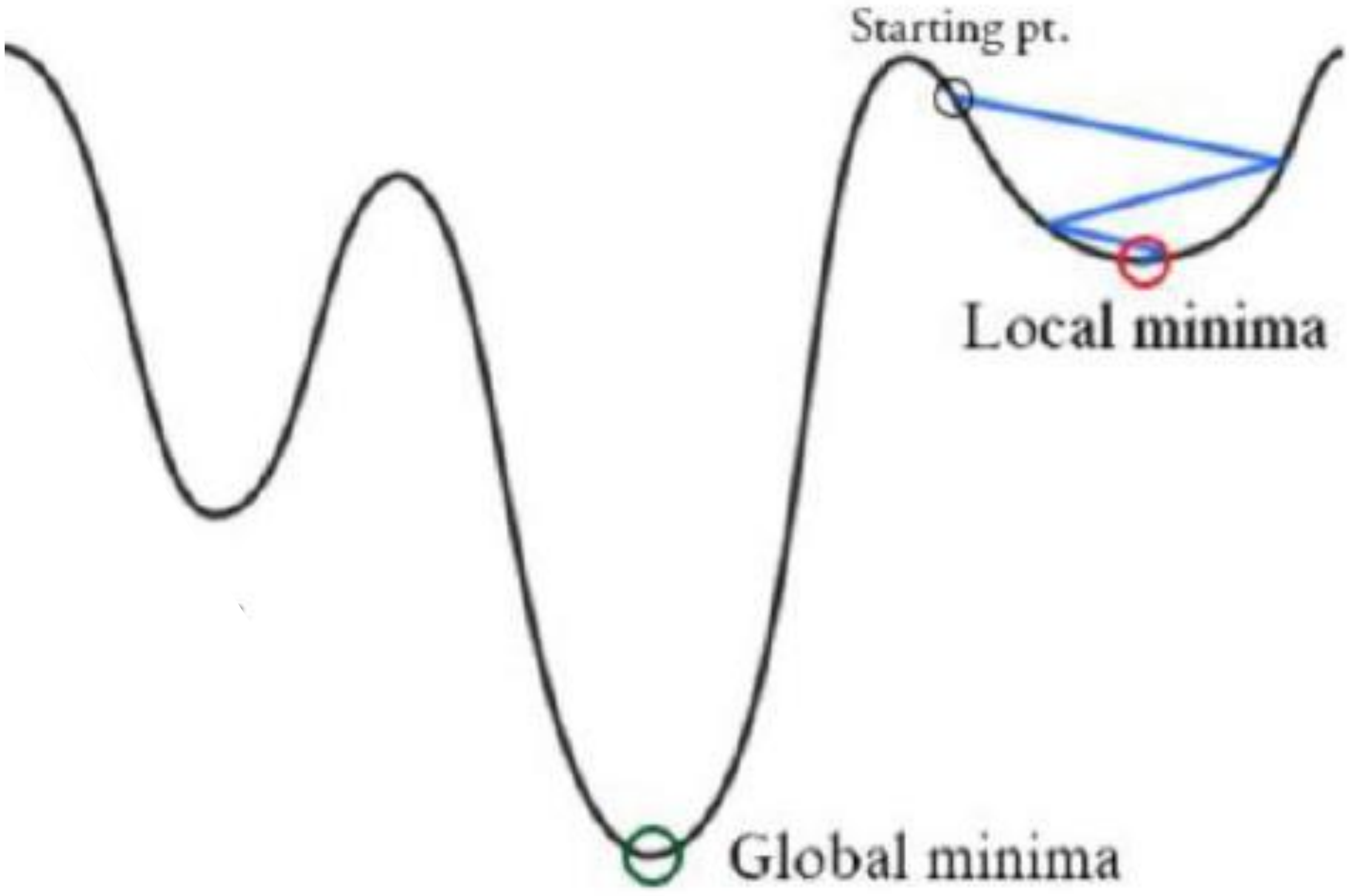


Image credit: https://www.cs.ubc.ca/labs/lci/mlrg/slides/non_convex_optimization.pdf

Non-Convex Optimization

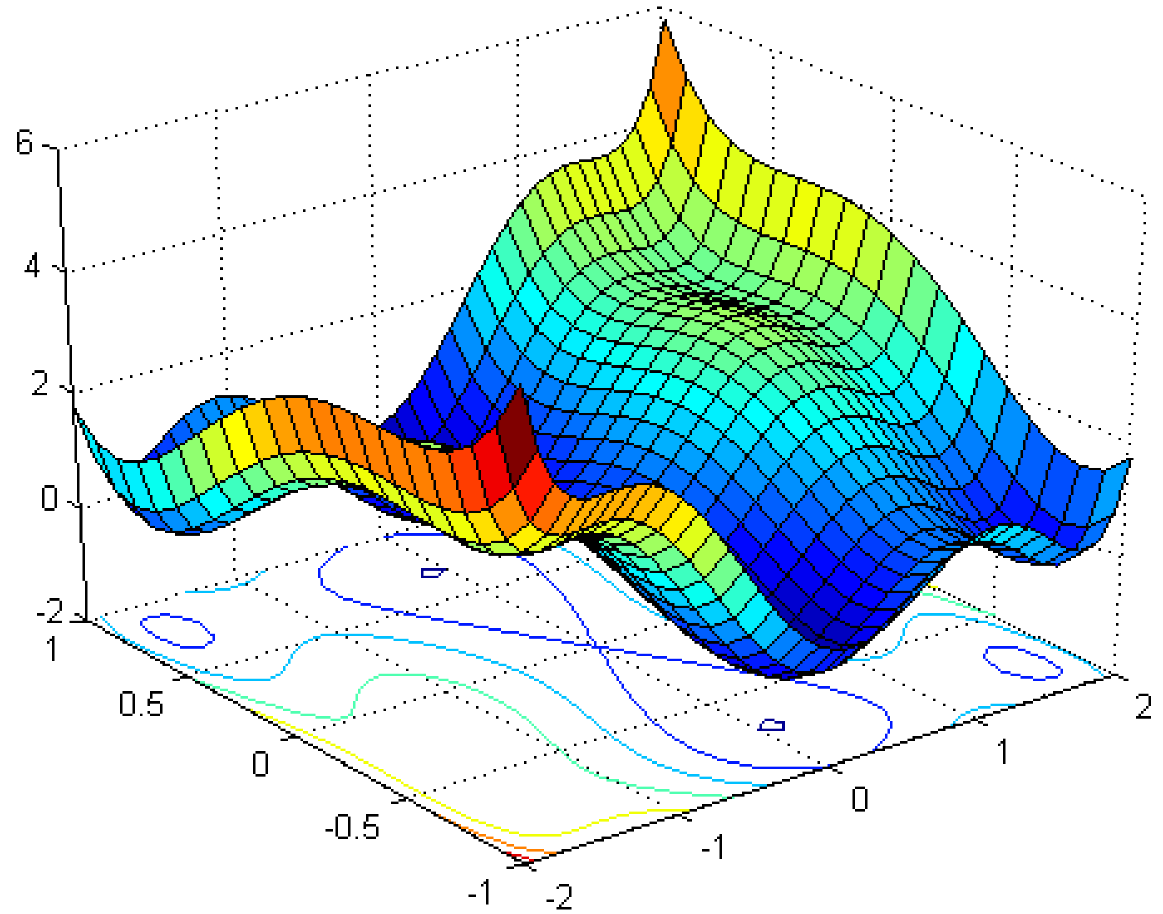


Image credit: https://stats.stackexchange.com/questions/279363/function-with-multiple-local-minima

Non-Convex Optimization

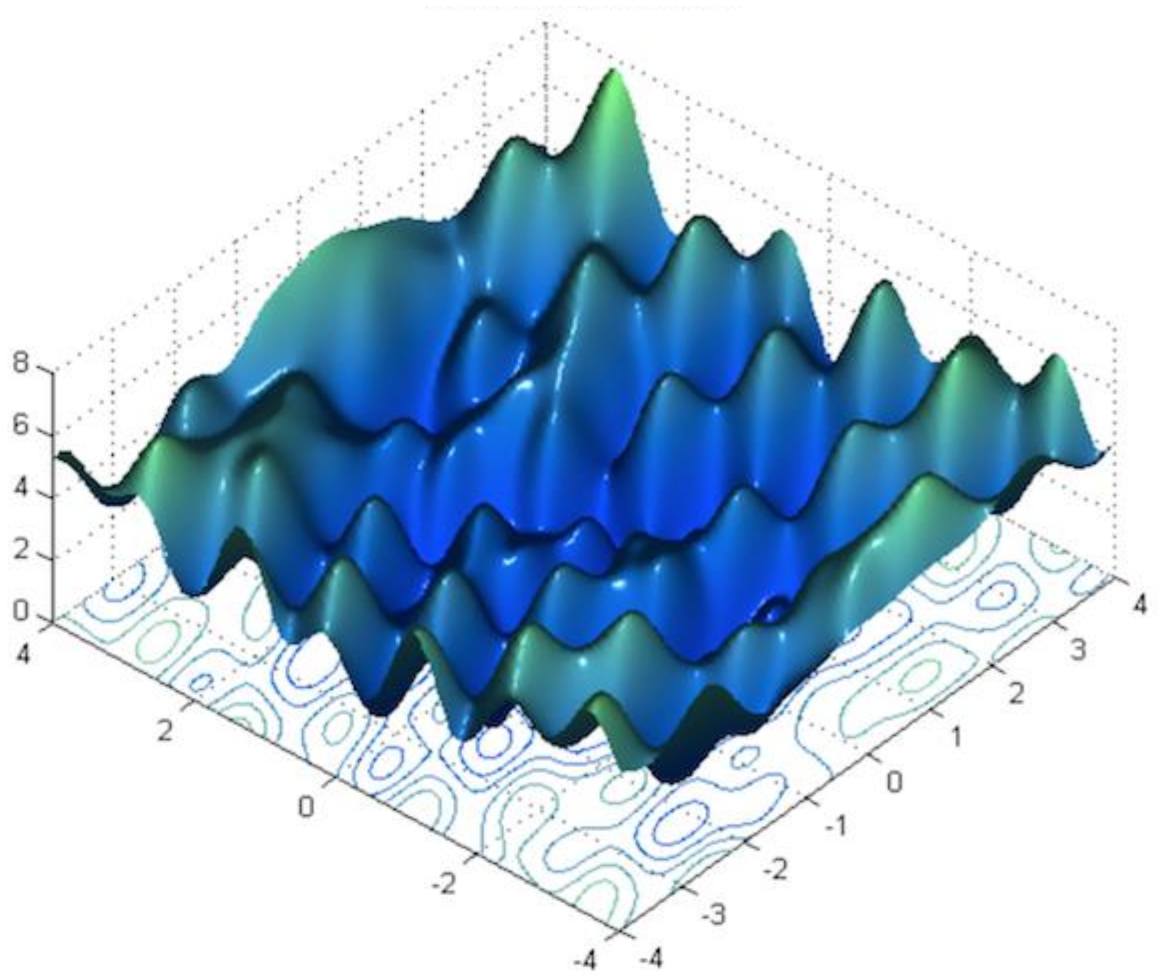


Image credit: https://towardsdatascience.com/neural-network-optimization-7ca72d4db3e0

Gradient Decent

Augustin-Louis Cauchy

An "iterative optimization algorithm for finding a local minimum of a differentiable function".

It is based of taking steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

Step size = slope (derivative or gradient) x "learning rate".

If we take steps proportional to the positive of the gradient, we approach a local maximum; this procedure is the **gradient ascent**.

Gradient descent was proposed by Cauchy in 1847.

Source: https://en.wikipedia.org/wiki/Gradient_descent

Image credit: https://en.wikipedia.org/wiki/Augustin-Louis_Cauchy

Gradient Decent (cont'd)

"Gradient descent is also known as **steepest descent**; but gradient descent should not be confused with the method of steepest descent for approximating integrals."

Source: https://en.wikipedia.org/wiki/Gradient_descent

Gradient Decent (cont’d)

Optimal solution

“Illustration of gradient descent on a series of level sets”

[Source: https://en.wikipedia.org/wiki/Gradient_descent](https://en.wikipedia.org/wiki/Gradient_descent)

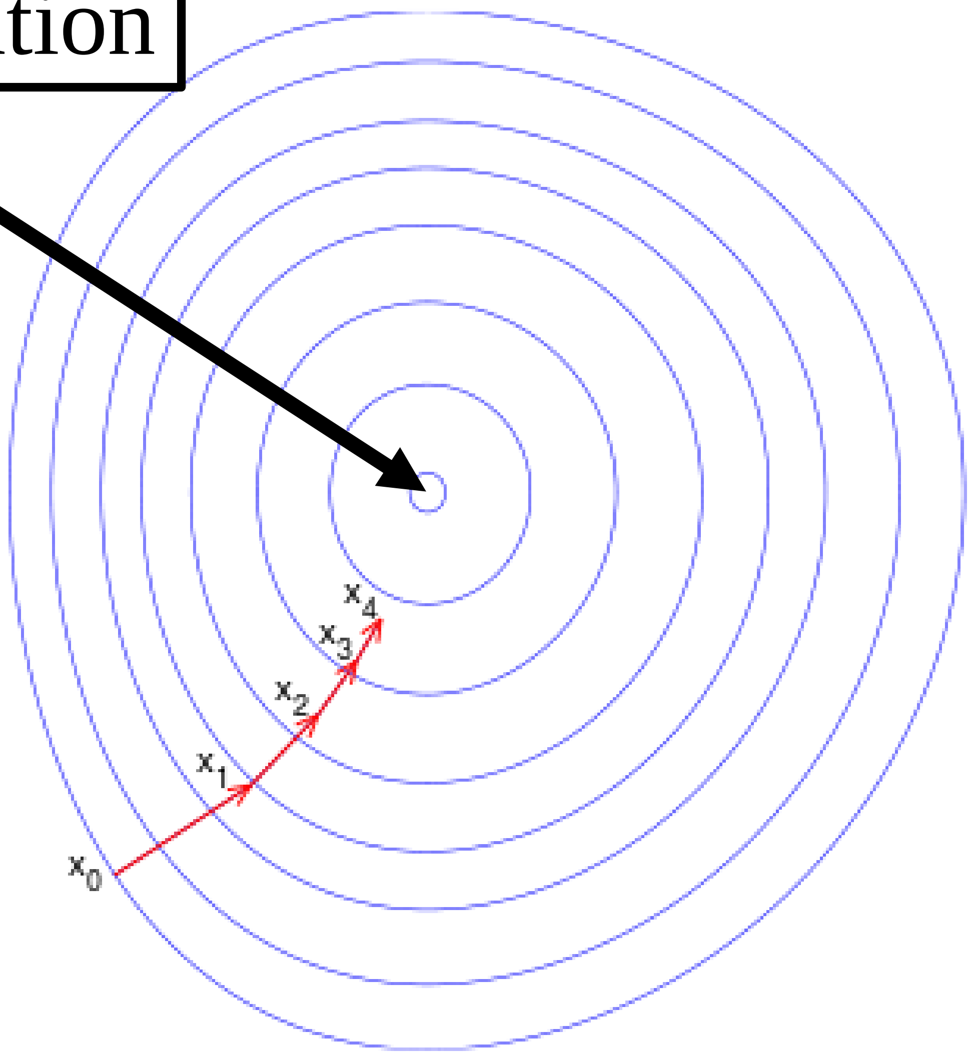


Image credit: https://upload.wikimedia.org/wikipedia/commons/thumb/f/ff/Gradient_descent.svg/700px-Gradient_descent.svg.png

Reducing Loss: Optimizing Learning Rate

https://developers.google.com/machine-learning/crash-course/fitter/graph

Other Links on Learning Rate

https://en.wikipedia.org/wiki/Learning_rate

https://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/

Introduction to Machine Learning

https://developers.google.com/machine-learning/crash-course/ml-intro

Gradient Descent, Step-by-Step

https://www.youtube.com/watch?v=sDv4f4s2SB8

Regression

Regression is a method used in many disciplines that attempts to determine the relationship between one dependent variable (usually denoted by Y or y) and other variables (known as independent variables).

Credit: https://www.investopedia.com/terms/r/regression.asp

The dependent variable (y) is related to one independent variable.

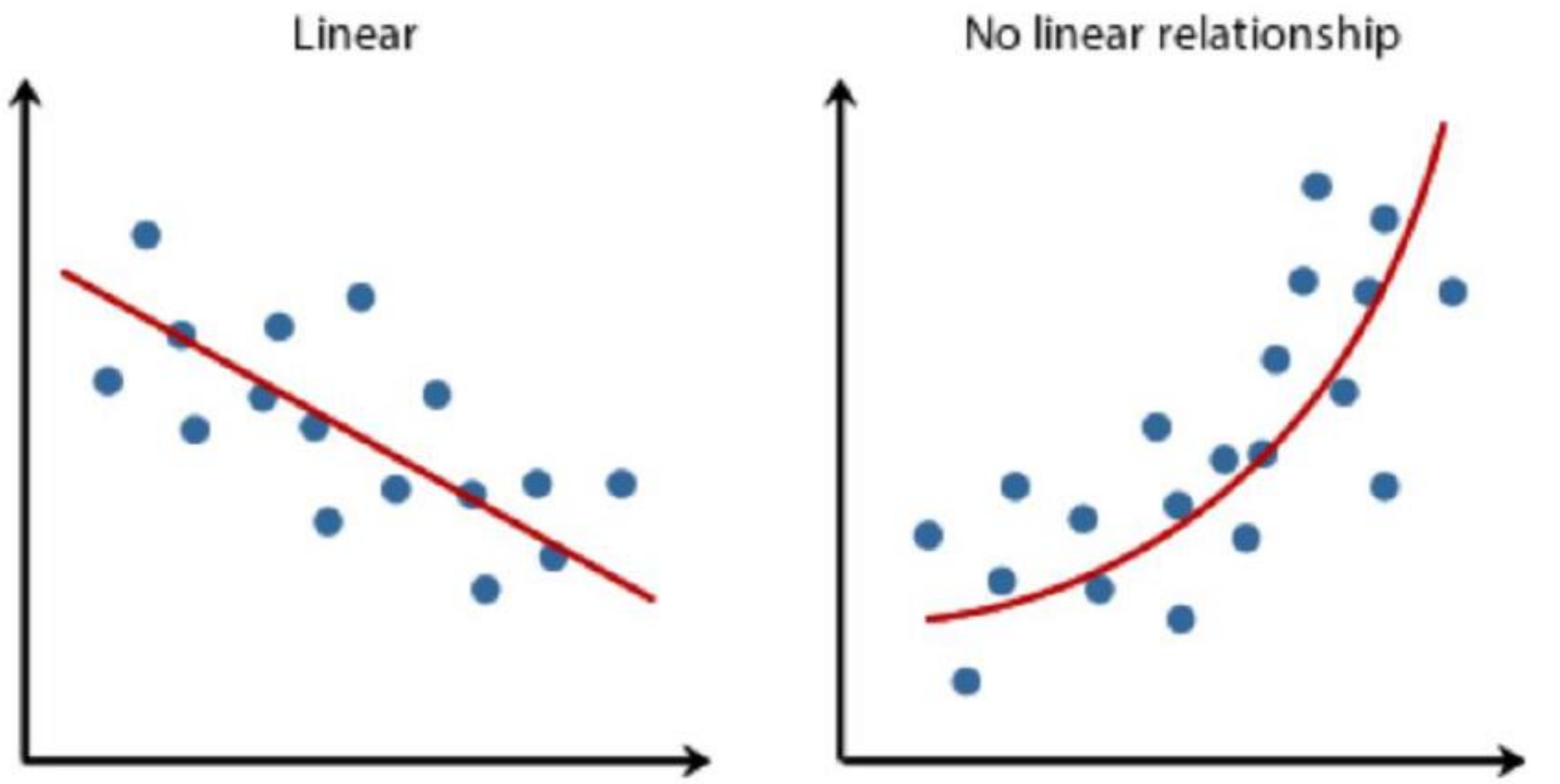
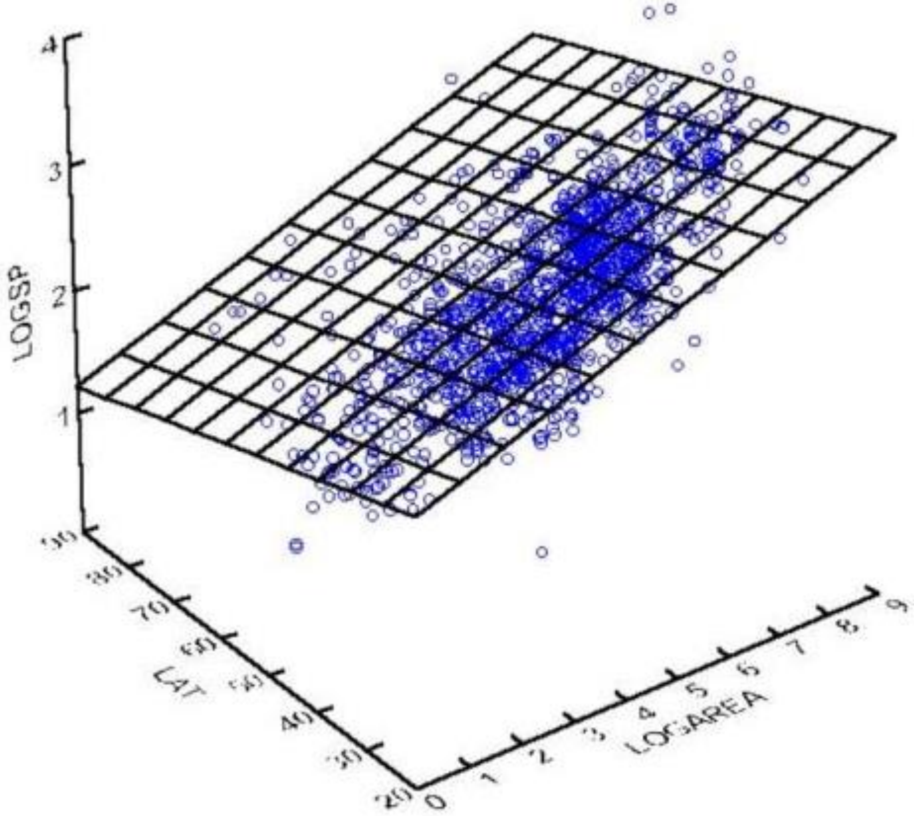


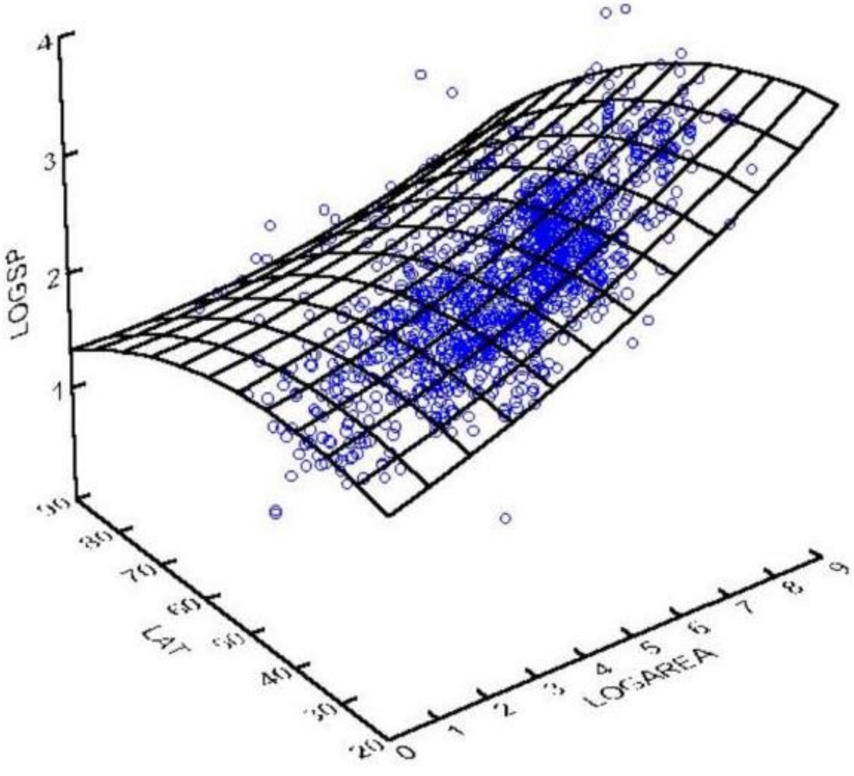
Image credit: https://hackernoon.com/linear-regression-and-its-mathematical-implementation-sc1ql24o8

The dependent variable (y) is related to more than one independent variable.

Linear regression



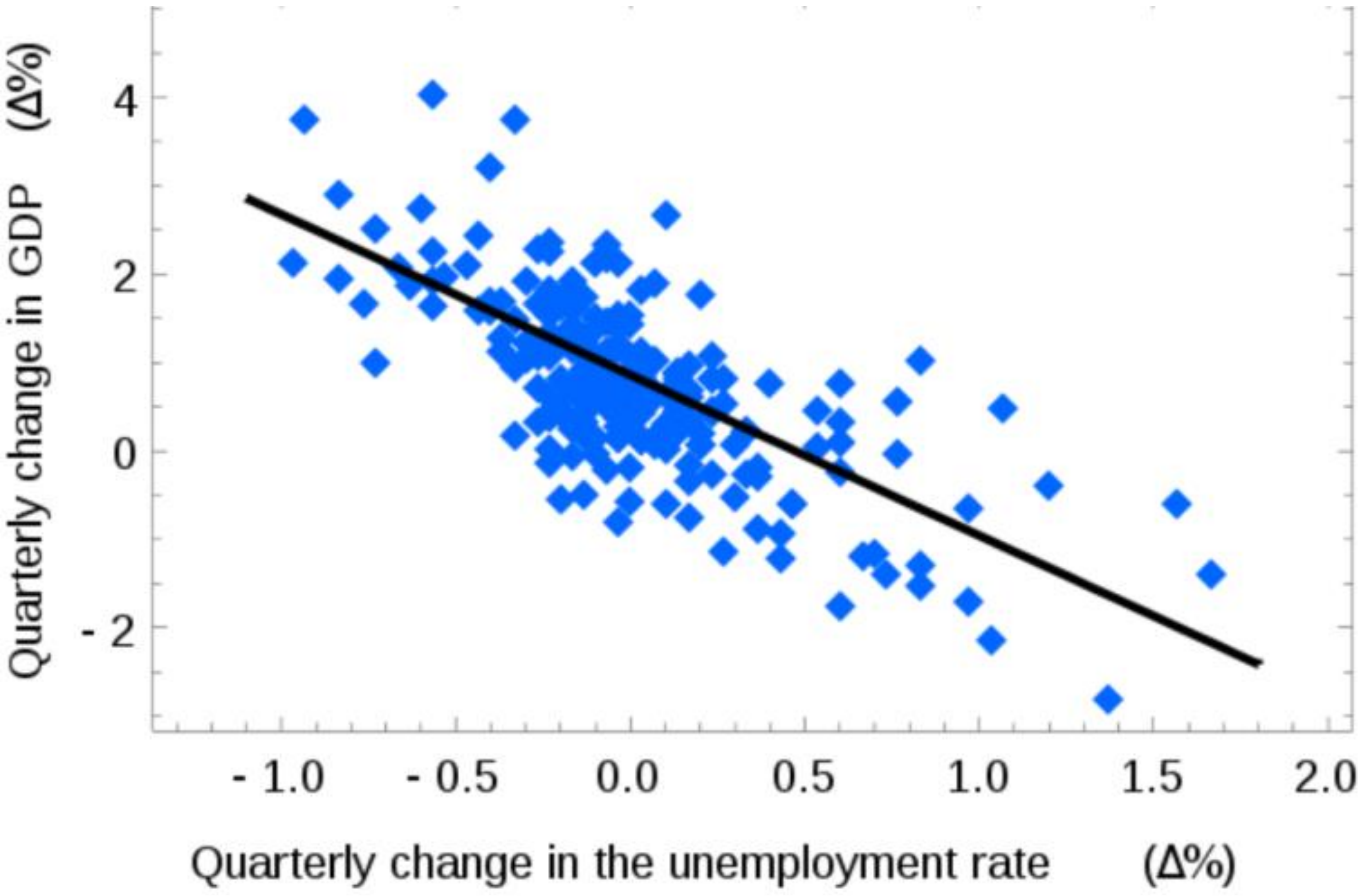
Non-linear regression



Credit: http://ordination.okstate.edu/MULTIPLE.htm

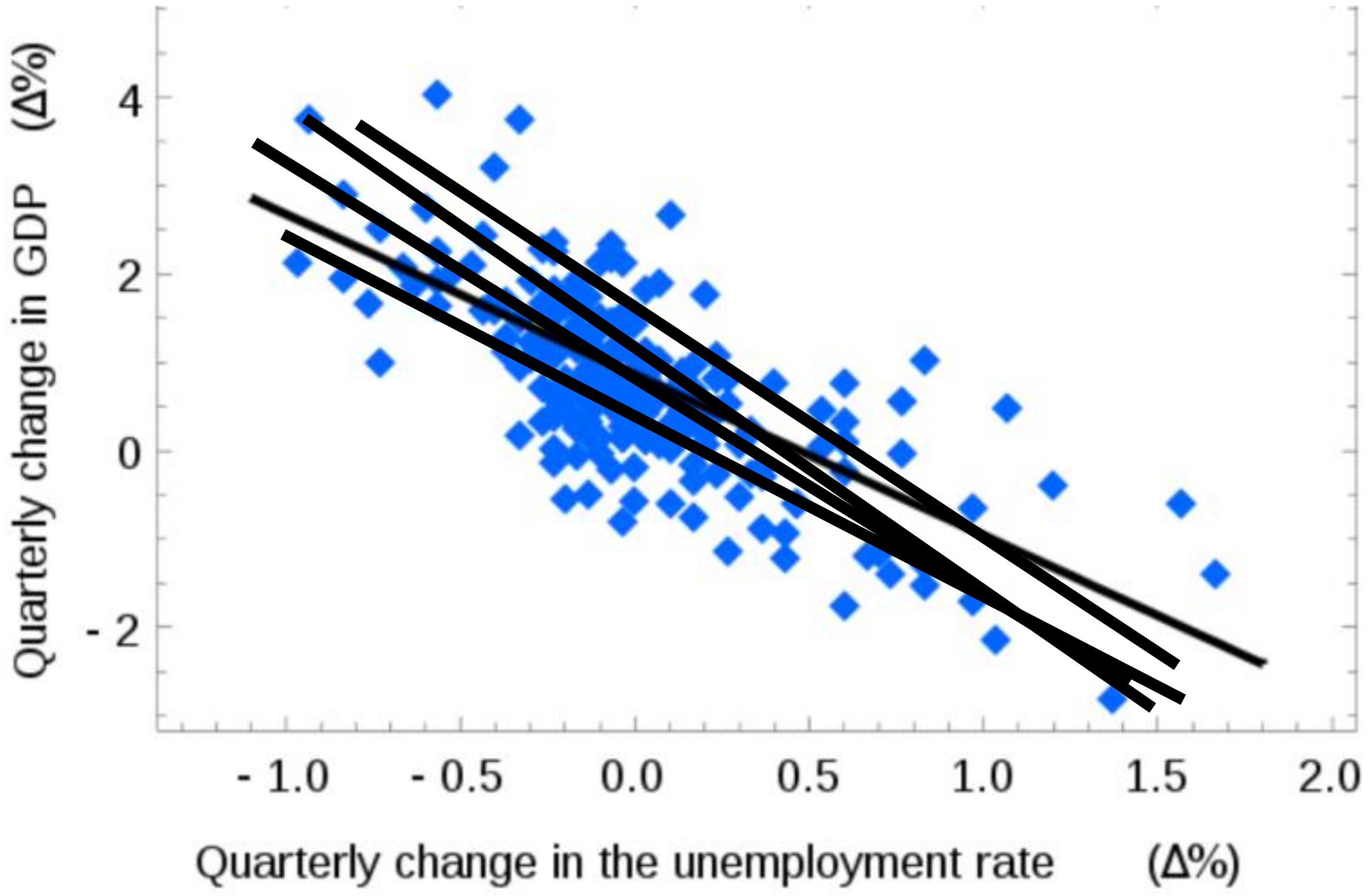
Linear Regression with One Dependent Variable Based on the Least-squares Approach

Okun's law in macroeconomics



Credit: Spasha - https://en.wikipedia.org/wiki/Simple_linear_regression

How to choose the line that is the best fit for the data?



[Credit: Spasha - https://en.wikipedia.org/wiki/Simple_linear_regression](https://en.wikipedia.org/wiki/Simple_linear_regression)

StatQuest Videos

https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw

Linear Regression

https://www.youtube.com/watch?v=nk2CQITm_eo

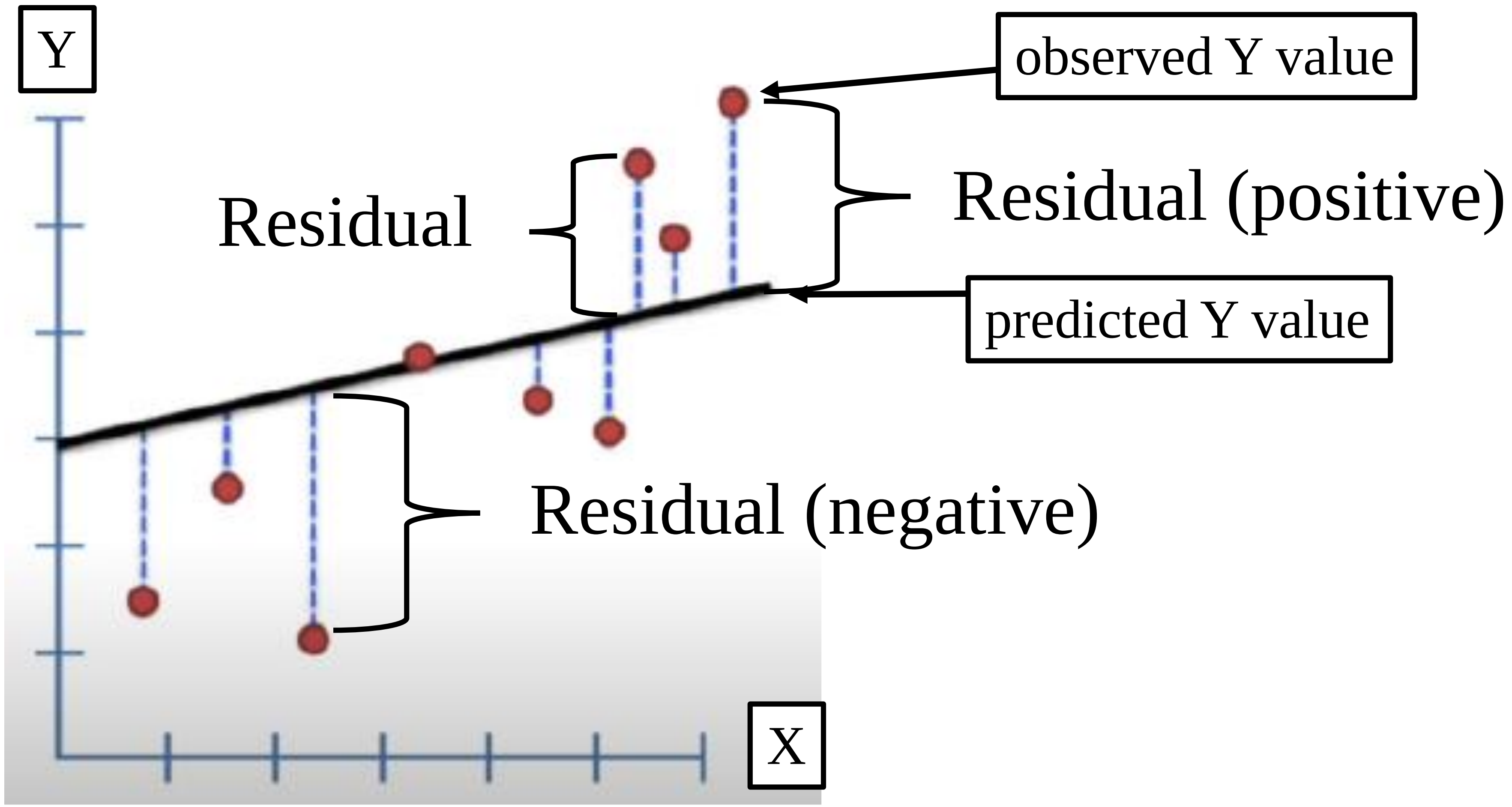
Multiple Regression

https://www.youtube.com/watch?v=zITIFTsivN8

Fitting a line to data, aka least squares, aka linear regression

https://www.youtube.com/watch?v=PaFPbb66DxQ

Residuals



Residual = [observed Y value] - [predicted Y value]

Image credit: https://www.youtube.com/watch?v=nk2CQITm_eo

Loss Function

Loss Function = Sum of the squares of the residuals
There are many loss functions that can be chosen, but the sum of the squares is popular.

<u>Why?</u>

1. **Summing up the residuals, positive residuals cancels out negative residuals.**
2. **Summing up absolute values, the resulted functions are not very amenable to analysis (e.g. not differentiable everywhere).**

Least Squares

The method of least squares is a standard approach for data fitting and regression analysis by minimizing the sum of the squares of the residuals to fit a curve (or a straight line) to a given set of data points.

Adrien-Marie Legendre (only portrait available)

Image credit: https://en.wikipedia.org/wiki/Adrien-Marie_Legendre

Carl Friedrich Gauss

Image credit: https://en.wikipedia.org/wiki/Least_squares

Least Squares (cont'd)

$(x_i, y_i),\ i = 1, 2, 3, \dots,\ n$, is a set of given data points where x_i, is an independent variable and y_i is a dependent variable.

$f(x, \boldsymbol{p})$ is the model function, where x represents an independent variable and $\boldsymbol{p}$ is a vector of parameters that we need to tune so that $f(x, \boldsymbol{p})$ is a best-fit predictor of y which is the dependent variable corresponding to the independent variable x.

The fit of a model to a data point is measured by its residual: $r_i = y_i - f(x_i, \boldsymbol{p})$.

Source: https://en.wikipedia.org/wiki/Least_squares

Least Squares (cont'd)

The least-squares method finds the optimal parameter values by minimizing the sum, S , of the squared residuals:

$$S = \sum_{i=1}^{n} (r_i)^2$$

or

$$S = \sum_{i=1}^{n} \left(y_i - f(x_i, \boldsymbol{p})\right)^2.$$

Source: https://en.wikipedia.org/wiki/Least_squares

Least Squares for Simple Linear Regression

Consider the simplest case where we are interested to fit a straight line

$$y = ax + b$$

to a given set of data points: (x_i, y_i), $i = 1, 2, 3, \ldots, n$.
in this case, $y = f(x_i, \boldsymbol{p}) = ax + b$.

$$r_i = y_i - f(x_i, \boldsymbol{p}) = y_i - ax_i - b.$$

$$S = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

Credit: https://en.wikipedia.org/wiki/Simple_linear_regression

Least Squares for Simple Linear Regression (cont'd)

Now our aim is to find the best values of a and b that will minimize the convex function

$$S = \sum_{i=1}^{n} (y_i - \ a x_i - b)^2.$$

Credit: https://en.wikipedia.org/wiki/Simple_linear_regression

Least Squares for Simple Linear Regression (cont'd)

Define:

$\bar{x}$ = the average of all the x_i, namely, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$\bar{y}$ = the average of all the y_i, namely, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

Let $\hat{a}$ and $\hat{b}$ denote the optimal solutions for a and b, respectively.

Credit: https://en.wikipedia.org/wiki/Simple_linear_regression

Least Squares for Simple Linear Regression (cont'd)

Analytic Solution

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

Credit: https://en.wikipedia.org/wiki/Simple_linear_regression

Homework

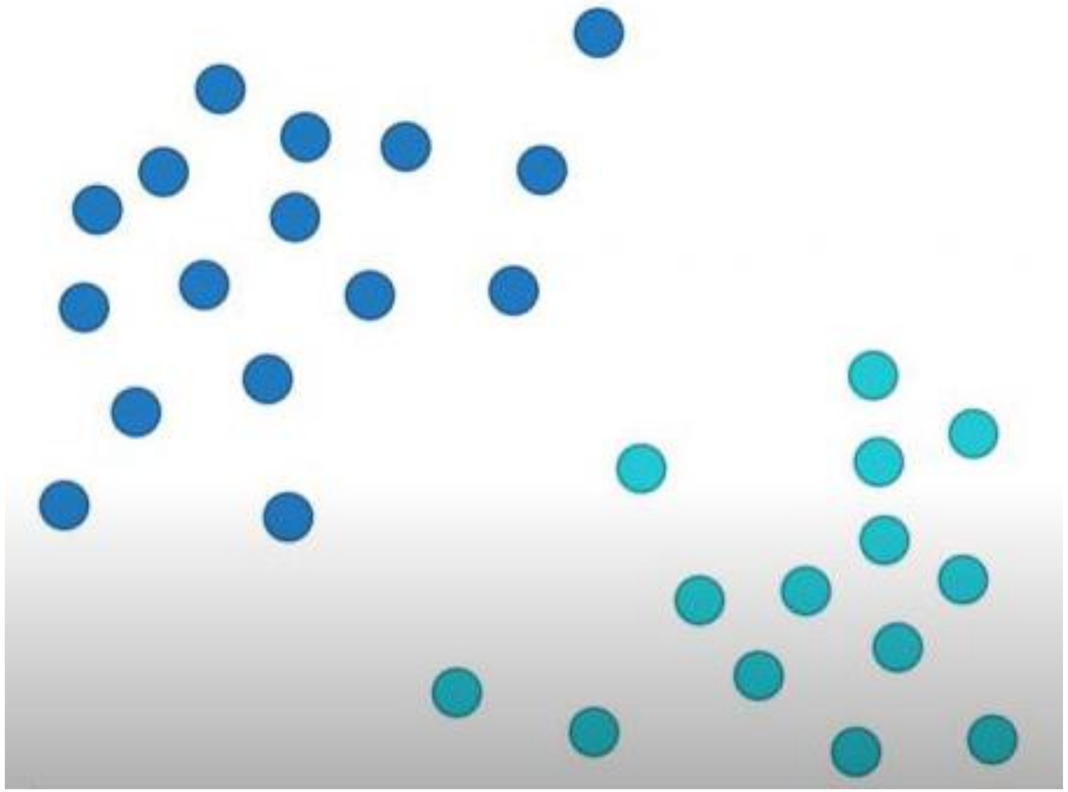
Apply the least squares method to fit a line to the following data.

Height (m), x_i	1.47	1.50	1.52	1.55	1.57	1.60	1.63	1.65	1.68	1.70	1.73	1.75	1.78	1.80	1.83
Mass (kg), y_i	52.21	53.12	54.48	55.84	57.20	58.57	59.93	61.29	63.11	64.47	66.28	68.10	69.92	72.19	74.46

Do it in the following three ways: (1) Use Excel to optimize the minimize S problem, (2) the analytical solution approach, and (3) Do regression in Excel (see https://www.youtube.com/watch?v=pEqKoRKZTwo). Compare the solutions and discuss discrepancies. Upload your solutions to Canvas/Discussions and discuss errors and discrepancies. (See Excel solution in the Excel file: convex_optimization/Sheet5)

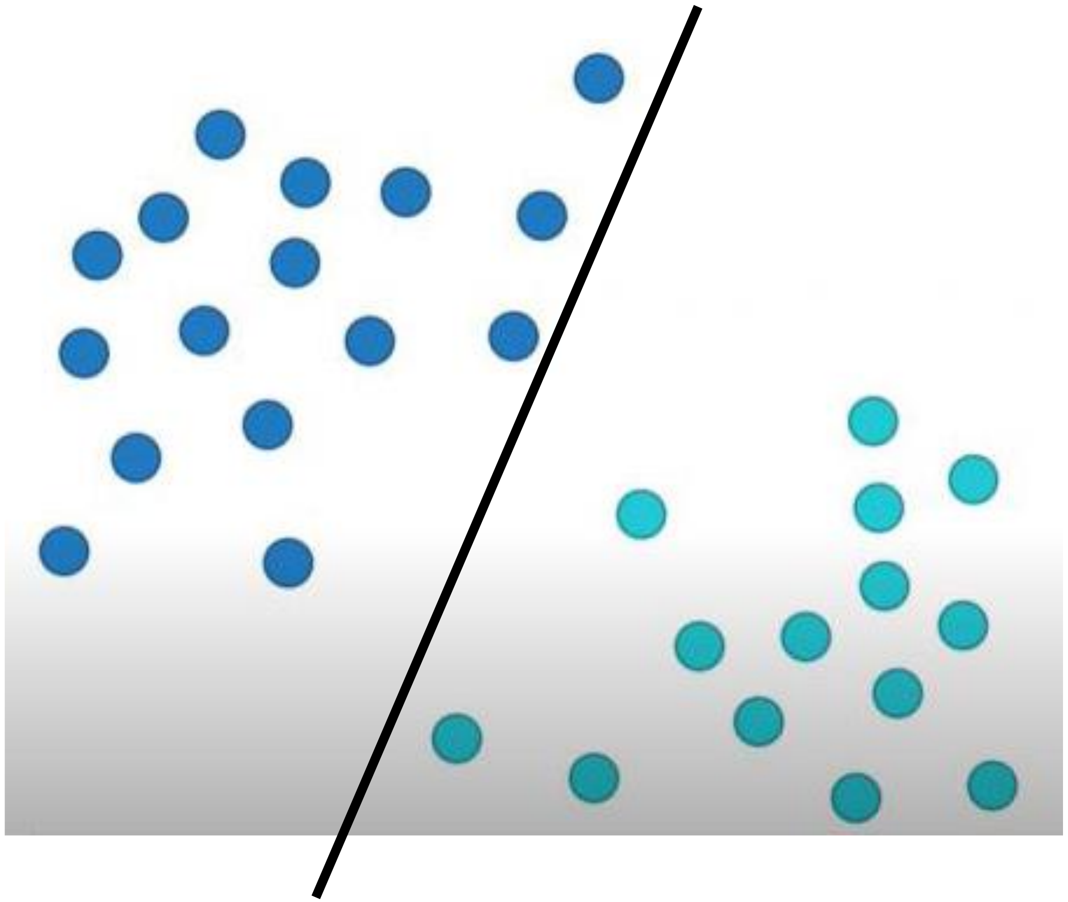
Source: https://en.wikipedia.org/wiki/Simple_linear_regression

Classification



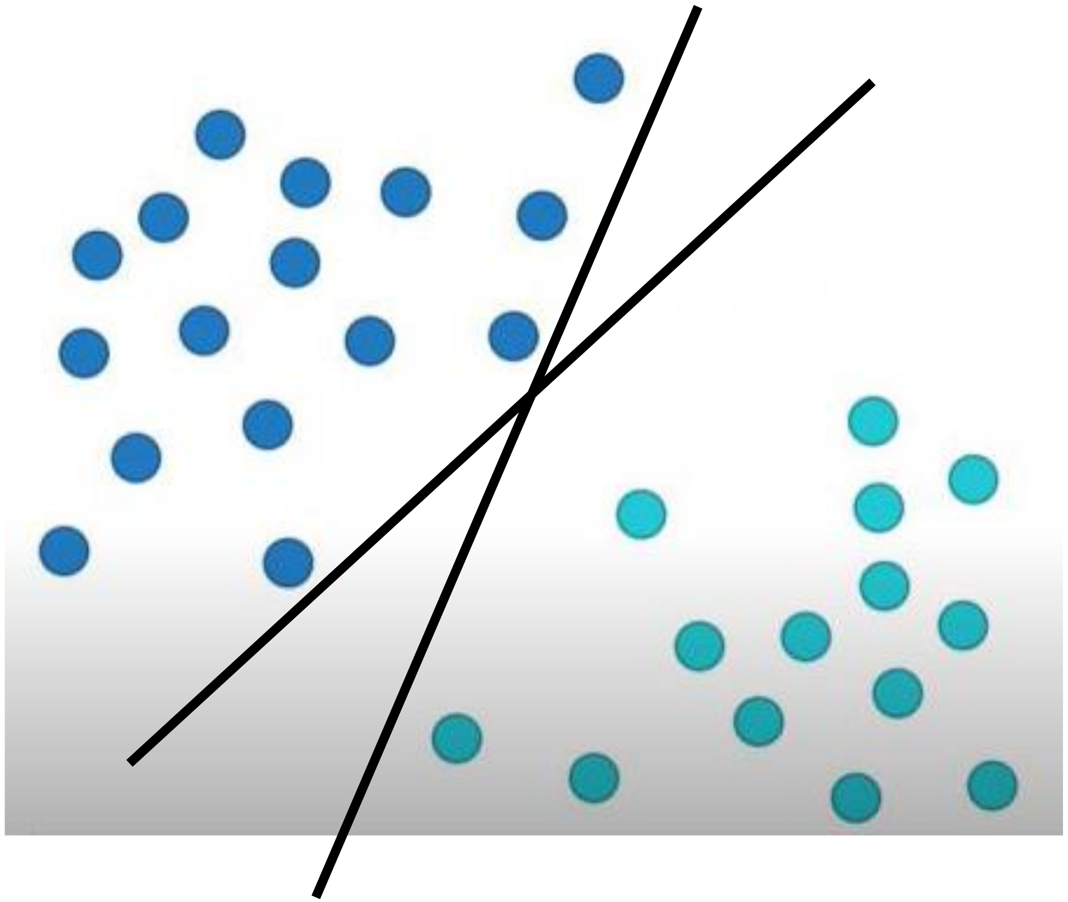
Source: https://www.youtube.com/watch?v=N1vOgolbjSc

Classification



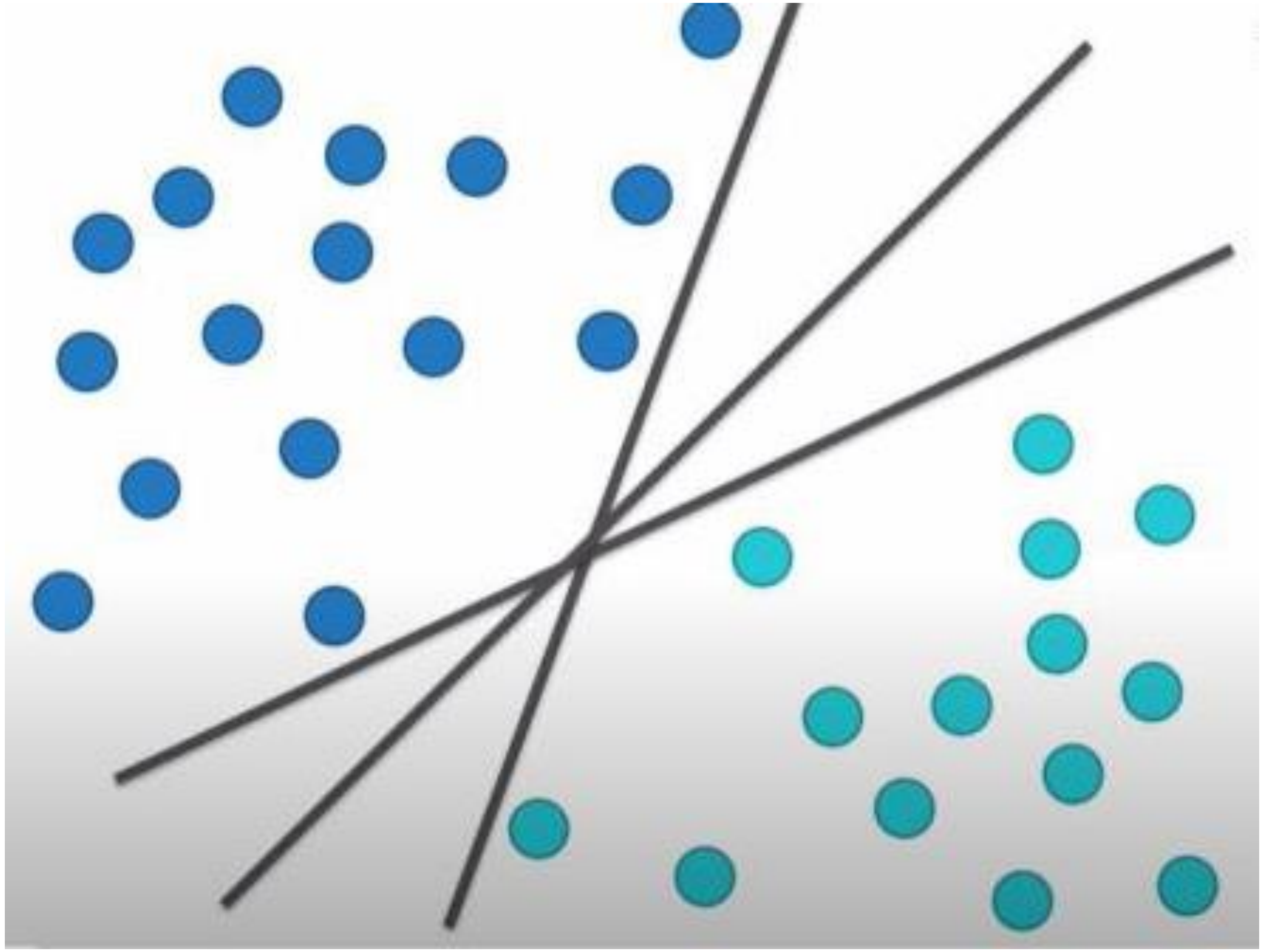
Source : https://www.youtube.com/watch?v=N1vOgolbjSc

Classification



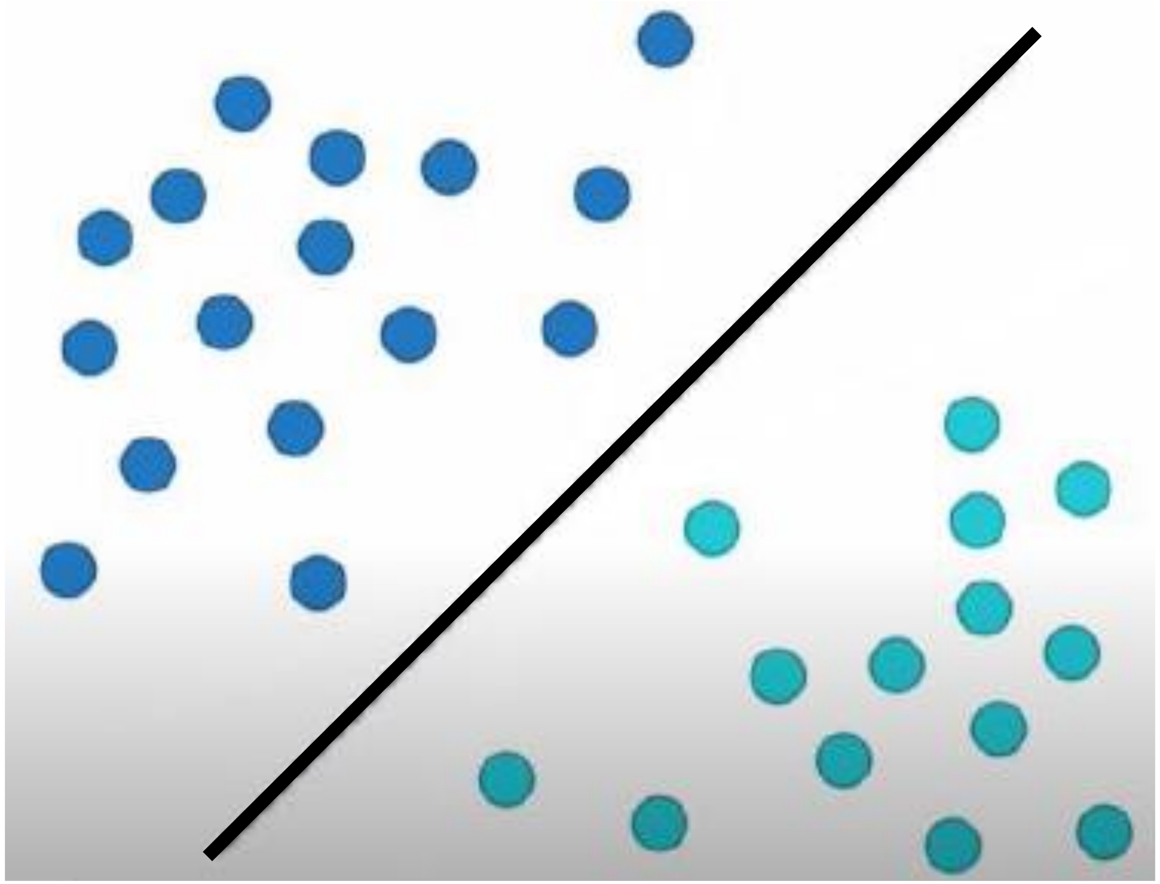
Source[: https://www.youtube.com/watch?v=N1vOgolbjSc](https://www.youtube.com/watch?v=N1vOgolbjSc)

Which line of the three is best?



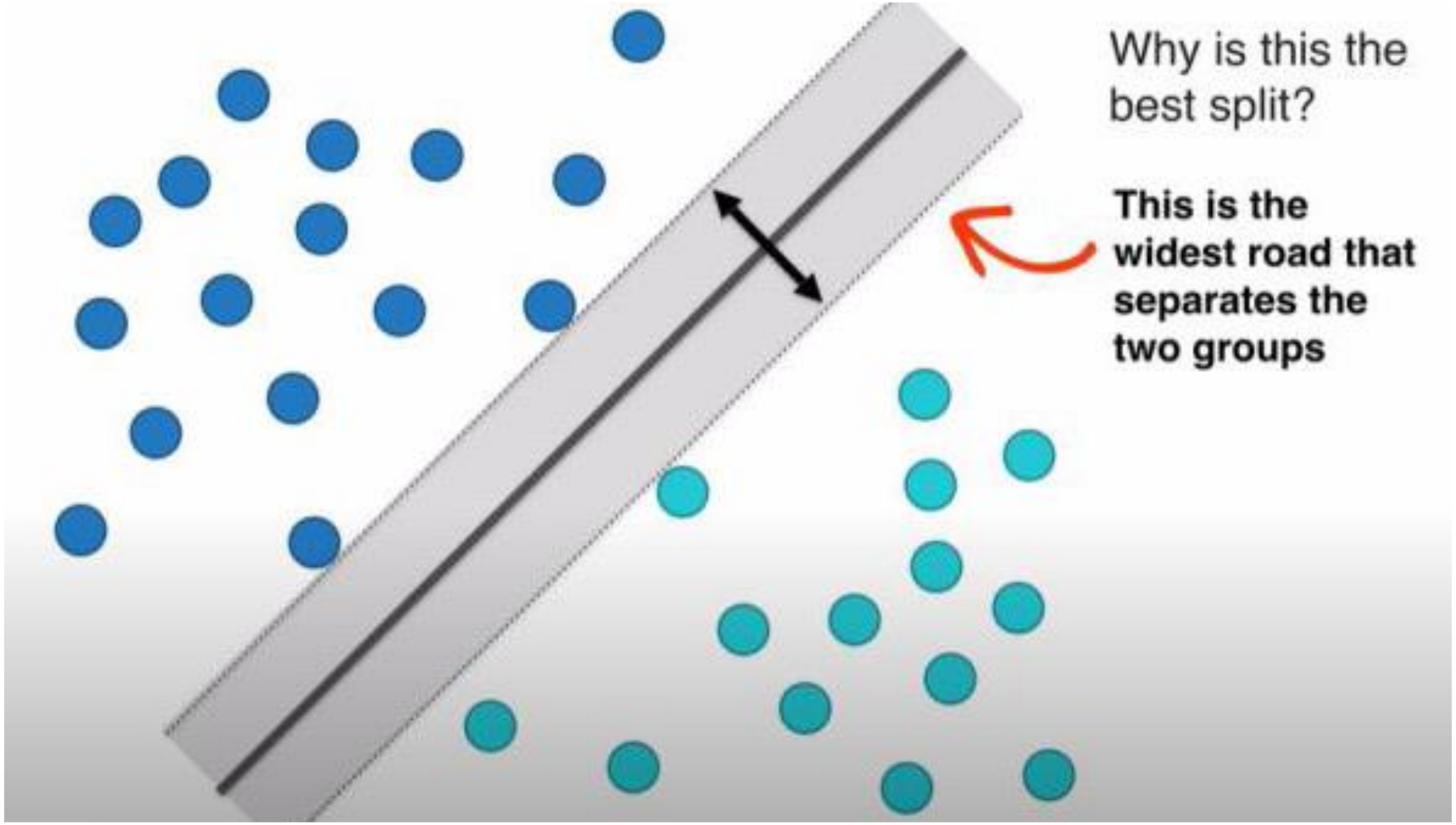
Source: https://www.youtube.com/watch?v=N1vOgolbjSc

Classification



Source: https://www.youtube.com/watch?v=N1vOgolbjSc

Classification (Support Vector Machine)



Source: https://www.youtube.com/watch?v=N1vOgolbjSc