Week 2. Convex Regression

May 25, 2024

1 Convexity

In some cases, we may believe that a curve we want to estimate is *convex.* A differentiable curve m is convex if its derivative is increasing.[1] More generally, a curve m is convex if all of its secants (line segments drawn from one point on the curve to another) lie above the curve, i.e., if for all points a and b it satisfies

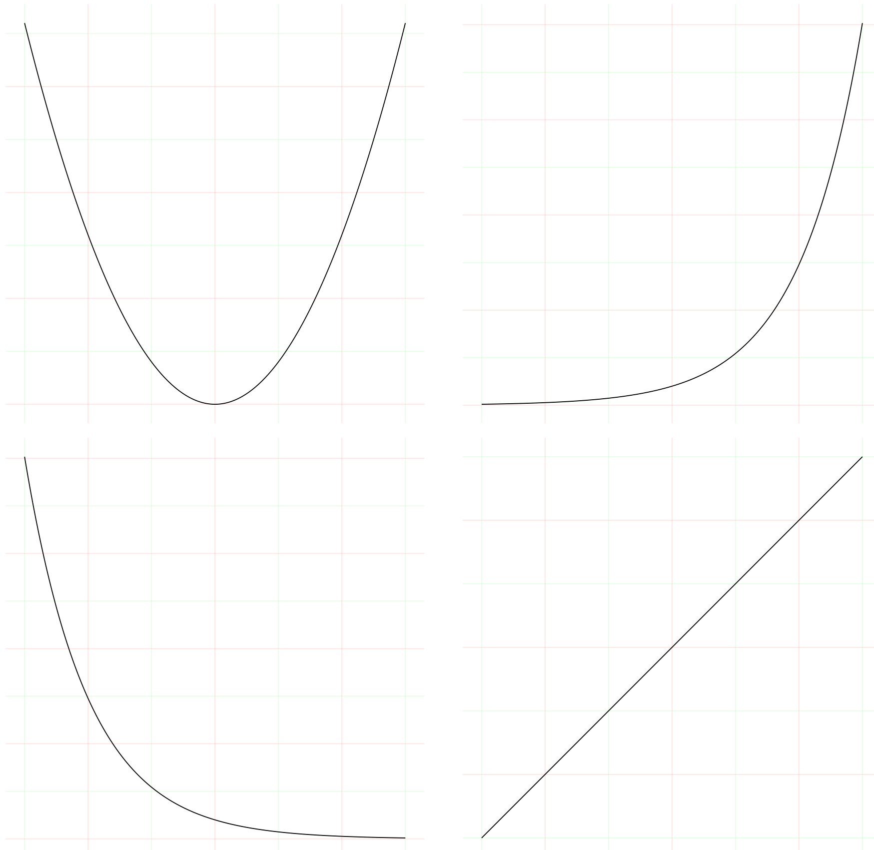
$$m\left\{(1-\lambda)a+\lambda b\right\} \le (1-\lambda)m(a)+\lambda m(b) \quad \text{for all} \quad \lambda \in [0,1]. \tag{1}$$

This inequality is, in mathematical notation instead of visual language, exactly what we said about secants. The left side is the height of the curve at $x_\lambda = (1-\lambda)a+\lambda b$ and the right is the height of the secant connecting a to b at x_λ.

Characterizing points on secants. Keep in mind that any point x on the segment between a and b can be written in the form $x_\lambda = (1-\lambda)a+\lambda b$ for $\lambda \in [0,1]$. To do this, we simply solve the equation $x_\lambda = (1-\lambda)a+\lambda b$ for λ in terms of x, i.e., we take $\lambda = (x-a)/(b-a)$. This is pretty intuitive: λ is the fraction of the distance from a to b that we have to travel to get from a to x.

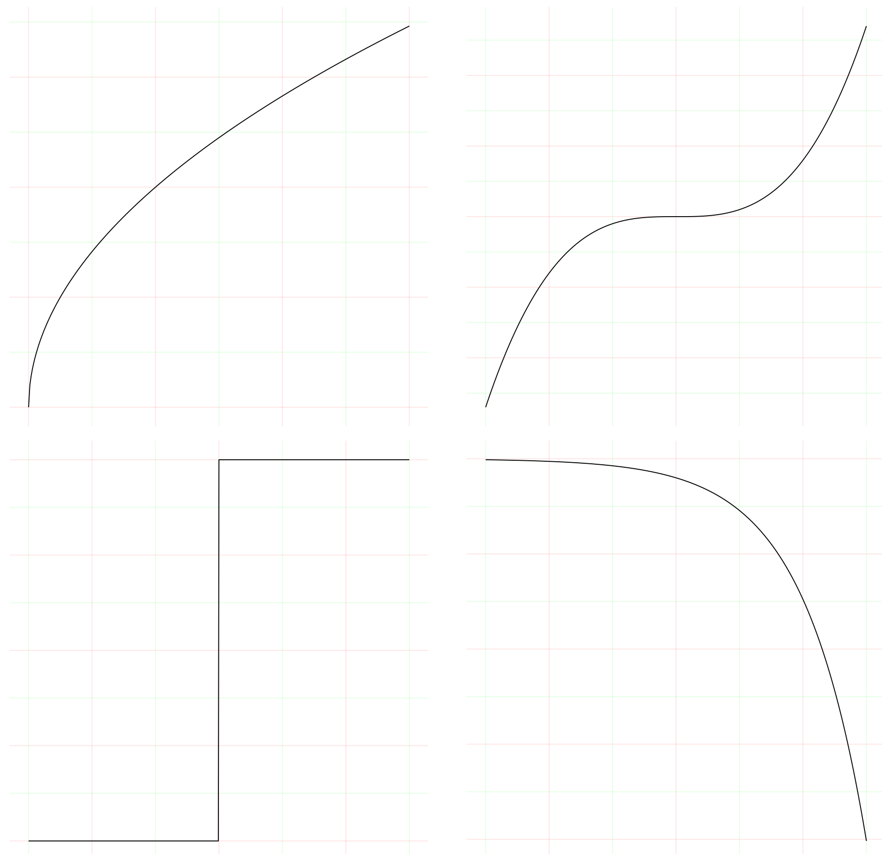
Examples. Here are some examples of convex curves.

1. $f(x) = x^2$
2. $f(x) = e^x$
3. $f(x) = e^{-x}$
4. $f(x) = x$



Here are a few curves that aren't convex.

1. $f(x) = \sqrt{x}$
2. $f(x) = x^3$
3. $f(x) = \begin{cases} 0 & \text{if } \; x < 0 \\ 1 & \text{if } \; x \geq 0 \end{cases}$
4. $f(x) = -e^x$



Exercise 1 *On the eight plots above, draw a few secants. For the non-convex curves, make sure at least one is below the curve somewhere between the secant's endpoints.*

1.1 Differentiable Convex Functions

Now that we've got a sense of what's going on visually, let's argue that our more general definition based on (1) is consistent with the informal definition based on derivatives I used in our first lecture.

Exercise 2 *Explain why, if a curve $m(x)$ is differentiable, it satisfies* (1) *if and only if its derivative $m'(x)$ is increasing.*

Hint. Here are two equivalent statements we can derive from (1) by taking $\lambda = (x-a)/(b-a)$ and $\lambda = 1-(x-b)/(b-a)$ respectively.

$$\begin{aligned} m(x) &\le m(a) + \frac{m(b)-m(a)}{b-a}(x-a) \qquad \text{for all } x \in [a,b] \\ m(x) &\le m(b) + \frac{m(b)-m(a)}{b-a}(x-b) \qquad \text{for all } x \in [a,b]. \end{aligned} \tag{2}$$

Rearranging, we get inequalities relating two slopes, one of which is the same in both cases.

$$\begin{aligned} \frac{m(x)-m(a)}{x-a} &\le \frac{m(b)-m(a)}{b-a} && \text{for all } x \in [a,b] \\ \frac{m(b)-m(a)}{b-a} &\le \frac{m(b)-m(x)}{b-x} = \frac{m(x)-m(b)}{x-b} && \text{for all } x \in [a,b]. \end{aligned} \tag{3}$$

What do these two equations together imply if we take $x \to a$ in the first and $x \to b$ in the second? This should help you show that convexity in the sense of (1) implies the increasingness of the derivative.

Another Hint. in the sense of (1). The mean value theorem tells us that, letting $x_\lambda = (1-\lambda)a + \lambda b$,

$$\begin{aligned} \frac{f(x_\lambda)-f(a)}{x_\lambda - a} &= f'(\tilde{a}) \qquad \text{for some point} \qquad \tilde{a} \in [a, x_\lambda] \\ \frac{f(b)-f(x_\lambda)}{b - x_\lambda} &= f'(\tilde{b}) \qquad \text{for some point} \qquad \tilde{b} \in [x_\lambda, b] \end{aligned} \tag{4}$$

If f' is increasing, how are these ratios related? And what, in terms of λ, a, and b, are their denominators? This should help you show that the increasingness of the derivative implies convexity in the sense of (1).

1.2 Convex Sets

There's a related notion of a *convex set*. We won't be using this for convex regression part of this homework, but it'll come up in lecture soon.

A convex set is a set that contains all line segments between points in it. That is, a set $\mathcal{S}$ is convex if and only if, for all points $a, b \in \mathcal{S}$, $(1-\lambda)a + \lambda b \in \mathcal{S}$ for all $\lambda \in [0,1]$. Here are a few examples.

In $1D$. A point, a line segment, or a line.

In $2D$. A filled-in triangle, square, or circle; the positive half-plane $\{(x,y) \in \mathbb{R}^2 : y \ge 0\}$; or the whole of $\mathbb{R}^2$.

Generally. A ball, the set $\{v : \rho(v) \le r\}$, of any radius r in any seminorm ρ.

Here are a few sets that aren't convex.

In $1D$. Two points. Or the union of two disconnected intervals, e.g. $\{x : x \in [-1, 0] \text{ or } [1, 2]\}$.

In $2D$. A not-filled-in triangle, square, or circle.

In $3D$. A sphere, the set $\{v : \|v\| = r\}$, of any radius $r > 0$ in any norm.

Exercise 3 *Prove that a ball in a seminorm ρ is convex.*

Tip. Use the triangle inequality.

Exercise 4 *Using the norms we discussed in our Vector Spaces Homework, explain why that implies that the filled-in unit square $\{(x, y) : |x| \leq 1, |y| \leq 1\}$ and circle $\{(x, y) : x^2 + y^2 \leq 1\}$ are convex. Finally, draw the set $\{v \in \mathbb{R}^2 : \|v\|_1 \leq 1\}$.*

Exercise 5 *Prove that a sphere of nonzero radius in any norm is not convex.*

Tip. Revisit the proof that seminorms are positive from the Vector Spaces Homework.

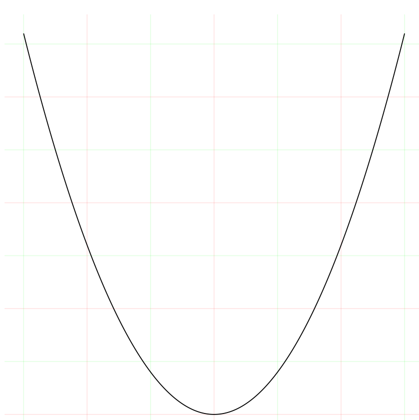
Exercise 6 *Draw, in 2D, a non-convex set that isn't included in the examples above.*

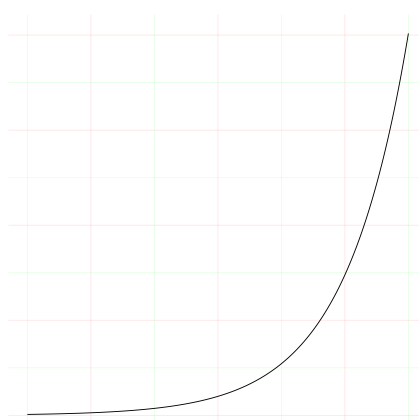
Exercise 7 *Explain why the intersection of two convex sets, i.e. the set of points that are in both of them, is a convex set.*

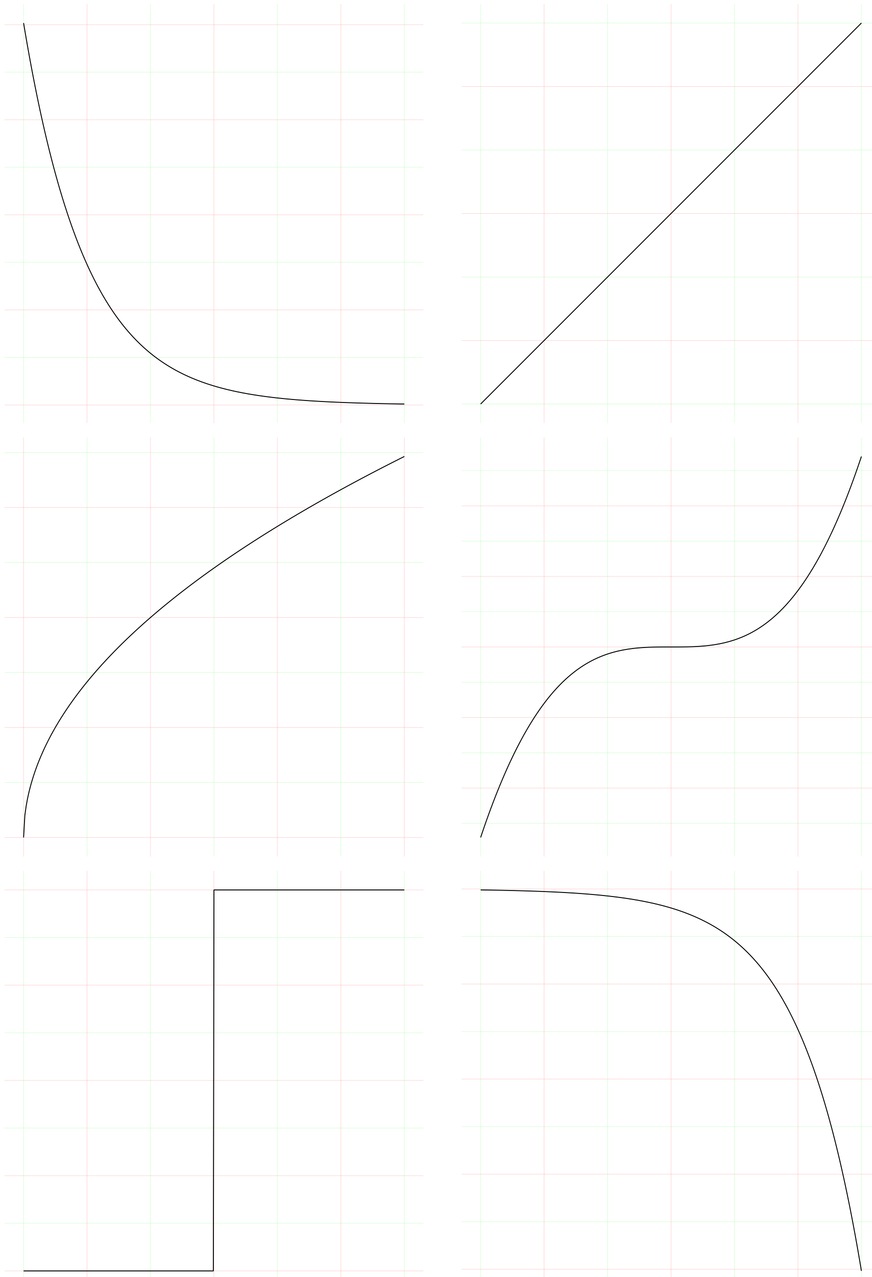
1.3 Convex functions have convex epigraphs

Here's another way of thinking about what convex functions look like. A function is convex if and only if its *epigraph*, the set of points on or above the curve, is convex. This is the definition of the epigraph of a function in mathematical notation. $\text{Epi}(f) = \{(x, y) : y \geq f(x)\}$.

Exercise 8 *On the plots below, fill in the epigraph.*







Exercise 9 *(Optional) Explain why this epigraph-based definition is equivalent to the secant-based definition above in* (1). *You don't have to give a formal proof.*

Tip. To show these definitions are equivalent, show that the convexity of a function's epigraph implies the convexity of the function and that the convexity of a function implies convexity of its epigraph. The latter part is a little harder. For intuition, try drawing a segment in a convex function's epigraph and the secant below it.

1.4 Thinking locally about convexity

Let's think about whether we can use a *local properties* to determine whether a function is convex. By local property, I mean something you can check by looking only at small pieces of the function rather than the whole function all at once. For example, we know a function is increasing everywhere if it's increasing between n and $n+1$ for all integers n. This works more generally, if in place of the intervals $[n, n+1]$ we use any set of intervals that combine to cover the whole real line. And because a differentiable function is convex if and only if it has an increasing derivative, it follows that we can use this approach to determine whether a differentiable function is convex.

Let's try to generalize this. To start, it's worth observing that using exactly this approach won't work.

Exercise 10 *Describe a non-convex curve that is convex on the intervals* $[n, n+1]$ *for all integers* n. *Here, by convex on an interval, I mean that* (1) *holds for all points* a, b *in that interval.*

Tip. Look at the examples of non-convex curves above.

We can fix this by looking at *overlapping intervals* that cover the real line, for example, the intervals $[n-1, n+1]$. By overlapping, I mean that the endpoints of each interval are in the interior of (i.e. in but not endpoints of) some other interval. Our ultimate goal will be to show that a function is convex if it's convex on overlapping intervals that cover the real line. But to get the concepts down without messy arithmetic, let's start with something easier.

Exercise 11 *Show that if* $f(1) \le \frac{1}{2}f(0) + \frac{1}{2}f(2)$ *and* $f(2) \le \frac{1}{2}f(1) + \frac{1}{2}f(3)$, *then* $f(1) \le \frac{2}{3}f(0) + \frac{1}{3}f(3)$. *Continue with this approach to show that* $f(1) \le \frac{3}{4}f(0) + \frac{1}{4}f(4)$ *if, in addition,* $f(3) \le \frac{1}{2}f(2) + \frac{1}{2}f(4)$.

It looks like there's a pattern here. If $f(n+1) \le f(n) + f(n+2)$ for positive integers n, then $f(1) \le \frac{n-1}{n}f(0) + \frac{1}{n}f(n)$. And because $1 = \frac{n-1}{n} \cdot 0 + \frac{1}{n} \cdot n$, this is an instance of our convexity-defining inequality (1) for $a = 0$ and $b = n$. If you're familiar with proof by induction, try the next exercise.

Exercise 12 *(Optional) Prove it! Use induction on* n.

The general case. If, for some increasing sequence $x_1 < x_2 < x_3 < \ldots < x_n$, a function f is convex on the overlapping intervals $[x_1, x_3]$, $[x_2, x_4]$, ..., $[x_{n-2}, x_n]$, then it's convex on the interval $[x_1, x_n]$. This is what we'll want when we're implementing convex regression.

Exercise 13 *(Optional) Prove it!*

Tip. Start by showing that if f is convex on two intervals $[a, b]$ and $[b, c]$ and satisfies $f(b) \le (1-\lambda')f(a) + \lambda' f(c)$ for the value of $\lambda' \in [0, 1]$ for which $b = (1-\lambda')a + \lambda' c$, then f is convex on $[a, c]$. To do this, it helps to observe that we can write $x \in [a, b]$ as $(1-\lambda)a + \lambda b = (1-\lambda)a + \lambda\{(1-\lambda')a + \lambda' c\} = (1-\lambda\lambda')a + \lambda\lambda' c$ and do something analogous for $x \in [b, c]$.

Tip. At some point in your argument, you'll probably want to take $a = x_1$, $b = x_3$, and $c = x_4$. To show that $f(x_3) \le (1-\lambda')f(x_1) + \lambda' f(x_4)$ for λ' such that $x_3 = (1-\lambda')x_1 + \lambda' x_4$, you'll want to use the properties that $f(x_3) \le (1-\lambda'')f(x_2) + \lambda'' f(x_4)$ for λ'' such that $x_3 = (1-\lambda'')x_2 + \lambda'' x_4$ and $f(x_2) \le (1-\lambda''')f(x_1) + \lambda''' f(x_3)$ for λ''' such that $x_2 = (1-\lambda''')x_1 + \lambda''' x_3$.

2 Convex Regression

Now that we've developed some intuition for what a convex function is, let's implement convex regression. That is, let's solve

$$\hat{\mu} = \underset{\text{convex } m}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \{Y_i - m(X_i)\}^2. \tag{5}$$

For this, we'll follow the same steps we used in the Monotone Regression Lab.

2.1 Fitting

We'll want to start with a notion of convexity we can implement as a constraint in CVXR. Think back to the Monotone Regression Lab, where we said a function m was *increasing on the sample* if $m(X_i) \le m(X_j)$ whenever $X_i \le X_j$. We used this to solve for the values $\hat{\mu}(X_1) \ldots \hat{\mu}(X_n)$ of the least squares solution on our sample by optimizing over an n-dimensional vector, then came up with a way of extending it into an outright increasing function that takes on these values so we could make predictions elsewhere.

We'll do the same here. We'll come up with a definition of *convexity on the sample* that, like increasingness on the sample, gives us constraints involving only the values $m(X_1) \ldots m(X_n)$. In the next section, we'll come up with a way of defining an outright convex function that takes on these values.

Exercise 14 *We'll say a function m is* convex on the sample *if it satisfies* (1) *for all combinations of a, b, and $(1-\lambda)a + \lambda b$ (with $\lambda \in [0, 1]$) that coincide with a set of three observations X_i, X_j, X_k. Express this set of constraints more concretely, as a set of constraints that does not involve λ.*

The downside of this, from an implementation perspective, is that it involve *a lot* of constraints. The number of constraints grows like n^3. We can fix this. Assume $X_1 \ldots X_n$ are sorted in increasing order. Because convexity on the overlapping intervals $[X_1, X_3] \ldots [X_{n-2}, X_n]$ implies convexity, we can replace this with the constraint that m is convex on the (sub)sample contained in each of these intervals.

Exercise 15 *Describe this smaller set of constraints.*

Once you've worked out the set of constraints, you should be ready to implement it. The best place to start, in my opinion, is the monotone regression code in the `increasingreg-fast-solution` block of the Monotone Regression Lab's solution.

Exercise 16 *Implement it! Then, to check that it's working, modify the Monotone Regression Lab's* `show-fit-on-sample` *block to use convex rather than monotone regression and take a look at the points $\hat{\mu}(X_1) \ldots \hat{\mu}(X_n)$ you've fit. Then modify the block to sample data around the other three curves μ from the Convergence Rates Lab and repeat. Include these four plots as your solution to this exercise. You don't need to submit code.*

2.2 Prediction

To make predictions, we'll need an algorithm for drawing a convex curve through a set of points $(X_1, \hat{\mu}(X_1)) \ldots (X_n, \hat{\mu}(X_n))$ where $\hat{\mu}$ is convex on the sample. As in the Monotone Regression Lab, we'll want to draw simple curves between neighboring points in our sample and extend our curve outside the range of our data in a sensible way. However, in this case we cannot use a piecewise-constant curve. After all, the step function is not convex.

Exercise 17 *Propose an algorithm. Use the same standard we used in lab. You don't need to provide pseudocode, but try to be precise and complete in your instructions. I should be able to stop anyone on campus, give them your plot of the points $(X_1, \hat{\mu}(X_1)) \ldots (X_n, \hat{\mu}(X_n))$ you fit earlier and your instructions, and get a drawing of the curve you intended.*

Now you should be ready to implement it.

Exercise 18 *Implement it, then plot a convex curve through the set of points $(X_1, \hat{\mu}(X_1)) \ldots (X_n, \hat{\mu}(X_n))$ you fit earlier for each of the four curves μ. Include your plots as your solution. You don't need to submit code.*

If you were going to share your convex regression code, you'd want to be confident that your algorithm works, i.e., that it gives you a convex curve.

Exercise 19 *Explain why your algorithm gives you a convex curve. You don't have to give a formal proof.*

Notes

[1]When I say a function f is increasing, what I really mean is that it doesn't ever decrease. That is, I mean that if $x \geq y$, then $f(x) \geq f(y)$. Some people say a function like this is *nondecreasing*. When they say f is increasing, they mean that if $x > y$, then $f(x) > f(y)$; people like me call such a function *strictly increasing*. Similarly, I'll say a curve is decreasing if it doesn't increase, where others might say it's nonincreasing, and call a decreasing curve

strictly decreasing if it doesn't have any flat spots, i.e., if $f(x) < f(y)$ whenever $x > y$. The advantage of what I do is that it's more direct: we're saying what the function does and not what it doesn't do. The disadvantage is that it's a bit less well-aligned with the meaning of increasing in colloquial language.