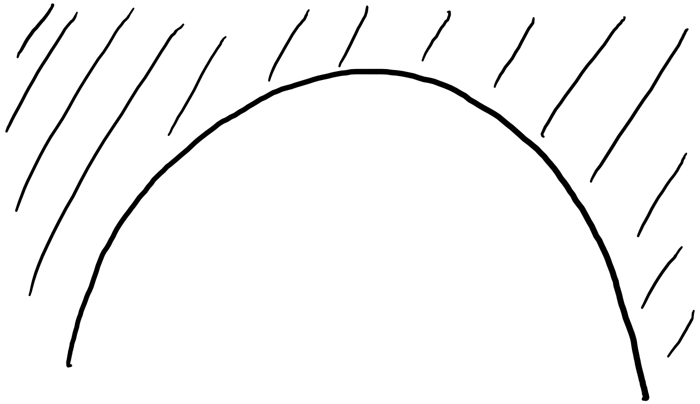
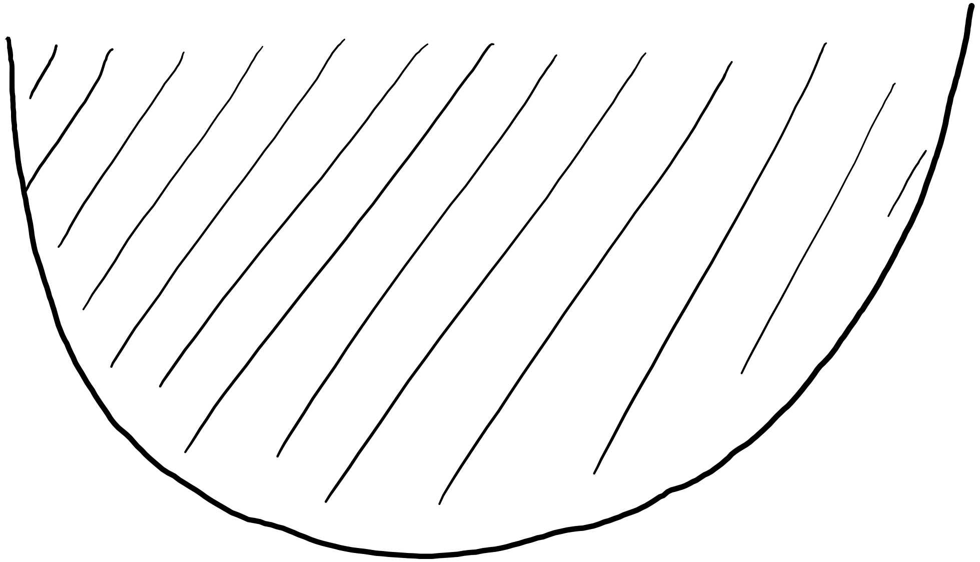
CPSC 340: Machine Learning and Data Mining

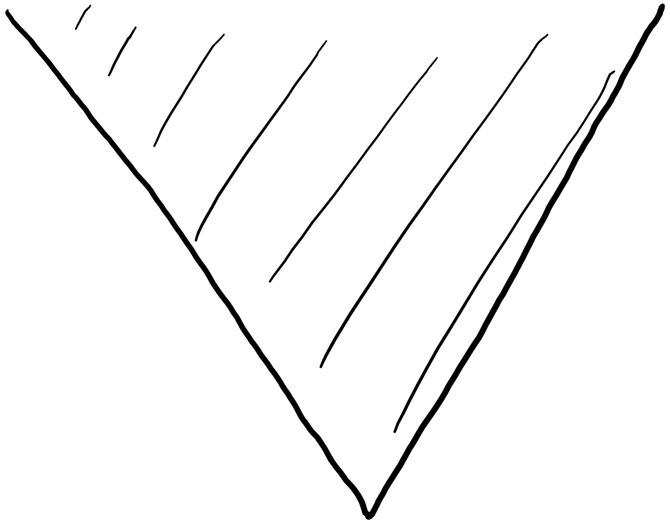
Convexity

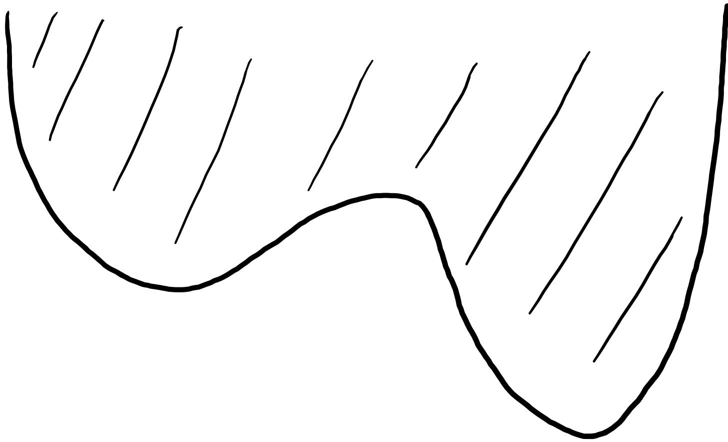
Convex Functions



- Is finding a 'w' with $\nabla f(w) = 0$ good enough?
  - Yes, for convex functions.







- A function is convex if the area above the function is a convex set.
  - All values between any two points above function stay above function.

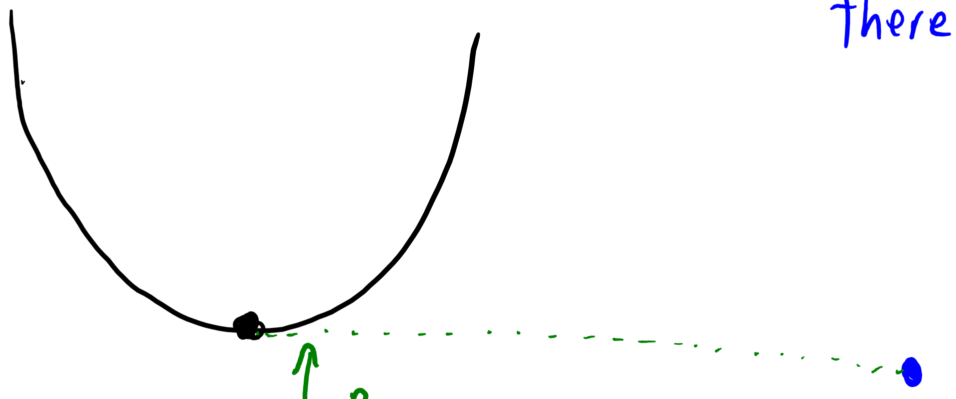
Convex Functions

- All ‘w’ with $\nabla f(w) = 0$ for convex functions are global minima.

Proof by contradiction:

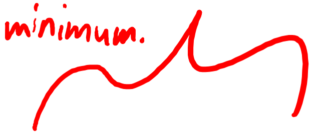
Consider a local minimum

If this is not global minimum, there must a smaller value.



By convexity we can move along line to global minimum and decrease objective.

But this contradicts that we are at a local minimum.



  – Normal equations find a global minimum because least squares is convex.

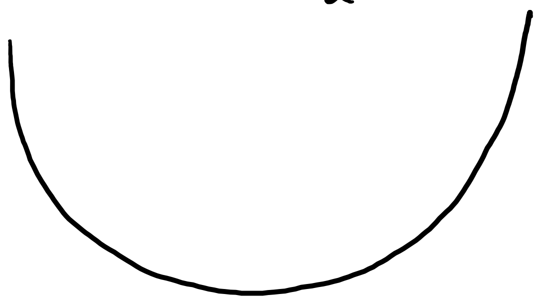
How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff f''(w) ≥ 0 for all 'w'.

Consider $f(w) = \frac{1}{2}aw^2$ for $a > 0$.

We have $f'(w) = aw$
and $f''(w) = a > 0$

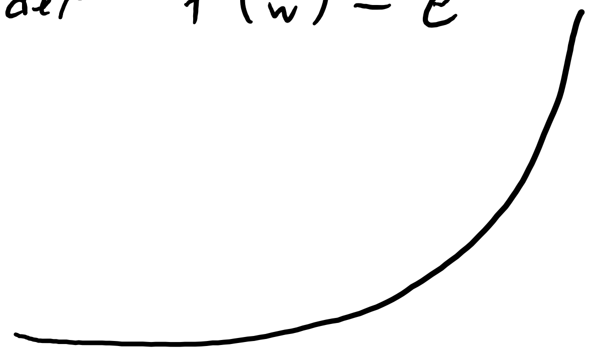
By assumption



Consider $f(w) = e^w$

We have $f'(w) = e^w$
and $f''(w) = e^w > 0$

By definition of exponential function.



How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.

We showed that $f(w) = e^w$ is convex, so $f(w) = 10e^w$ is convex.

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff f''(w) ≥ 0 for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.

$\|w\|$, $\|w\|^2$, $\|w\|_1$, $\|w\|_\infty$, $\|w\|_1^2$, and so on are all convex.

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.

$$f(w) = \underbrace{10e^w}_{\text{From earlier}} + \underbrace{\frac{\lambda}{2}}_{\text{constant}}\underbrace{\|w\|^2}_{\text{norm squared}} \quad \text{is convex}$$

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.

$$f(w) = w^T x_i = \underbrace{w_1 x_{i1}}_{\text{convex}} + \underbrace{w_2 x_{i2}}_{\text{convex}} + \cdots + \underbrace{w_d x_{id}}_{\text{convex}}$$

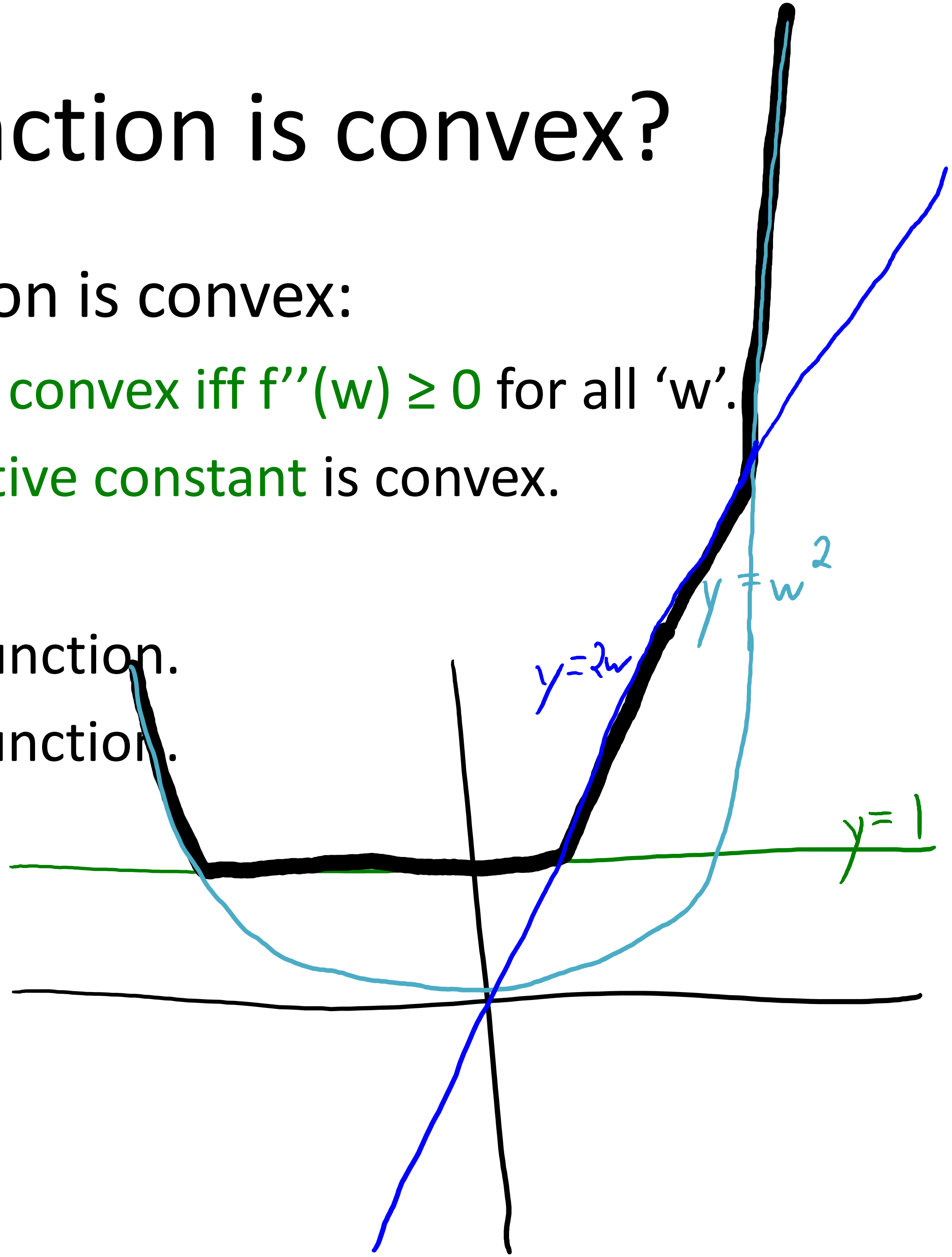
Second derivative of each term is 0.

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.
  - The max of convex functions is a convex function.

$f(w) = \max\{1, 2w, w^2\}$ is convex.

convex



How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.
  - The max of convex functions is a convex function.
  - Composition of a convex function and a linear function is convex.

If $f(w) = g(\underbrace{Xw - y}_{\text{linear function}})$ then 'f' is convex if 'g' is convex.

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.
  - The max of convex functions is a convex function.
  - Composition of a convex function and a linear function is convex.
- But: not true that multiplication of convex functions is convex:
  - If $f(x)=x$ (convex) and $g(x)=x^2$ (convex), $f(x)g(x) = x^3$ (not convex).

How do we know if a function is convex?

- Some useful tricks for showing a function is convex:
  - 1-variable, twice-differentiable function is convex iff $f''(w) \geq 0$ for all 'w'.
  - A convex function multiplied by non-negative constant is convex.
  - Norms and squared norms are convex.
  - The sum of convex functions is a convex function.
  - The max of convex functions is a convex function.
  - Composition of a convex function and a linear function is convex.
- Also not true that composition of convex with convex is convex:

Even if 'f' is convex and 'g' is convex, $f(g(w))$ might <u>not</u> be convex.

E.g., w^2 is convex and $(w-1)^2$ is convex, but $(w^2-1)^2$ is <u>not</u> convex.

Constraints, Continuity, Smoothness

- Sometimes we need to optimize with constraints:
  - Later we'll see "non-negative least squares".

$$\min_{w \geq 0} \frac{1}{2} \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

  - A vector 'w' satisfying $w \geq 0$ (element-wise) is said to be "feasible".
- Two factors affecting difficulty are continuity and smoothness.
  - Continuous functions tend to be easier than discontinuous functions.
  - Smooth/differentiable functions tend to be easier than non-smooth.
  - See the calculus review here if you haven't heard these words in a while.

Convexity, min, and argmin

- If a function is convex, then all critical points are global optima.

- However, convex functions don't necessarily have critical points:
  - For example, f(x) = a*x, f(x) = exp(x), etc.

- Also, more than one 'x' can achieve the global optimum:
  - For example, f(x) = c is minimized by any 'x'.