MATH 521 Lecture 18

SEQUENCES OF FUNCTIONS

Applied reason to care:

In machine learning, we are often trying to learn a function

$$f : \mathbb{R}^{100\times 100} \to [0,1] \qquad f(\text{image}) = \text{probability of dog}$$

$\mathbb{R}^{100\times100}$ = space of 100×100 images; $\{0,1\}$

The way you do this is by training:

You have labelled data $d_1, d_2, d_3, \vdots$ (image, dog / not dog)

f_n is the function the machine learns after seeing n data points

You will hope that, in some sense, that f_n converges to f as $n \to \infty$

More precisely, can ask if the f_n are in some sense Cauchy seen to be "settling down" to something

The distance between f_n and f is called the loss and we hope this is $\to 0$ as n gets large

Which notion of distance? L^∞ loss, L^2 loss, L^1 loss

L^∞ loss is big if there is even a single example where $f_n(x) = 0$ but $f(x) = 1$

L^2 loss is big if there are many modestly sized failures

Last time: $f_1, f_2, \ldots \quad X \to \mathbb{R}$

We said $f_1, f_2, \ldots$ converges pointwise to f

if, $\forall x \in X$, $\lim_{i\to\infty} f_i(x) = f(x)$

Def We say $f_1, f_2, \ldots$ converges uniformly to f if

for all $\varepsilon > 0$, $\exists N_\varepsilon$, s.t. $\forall x \in X$, $\forall n > N_\varepsilon$ $|f_n(x) - f(x)| < \varepsilon$

Rk Pointwise converges:

$\forall x \in X$, for all $\varepsilon > 0$, $\exists N_{x,\varepsilon}$ s.t. $\forall n > N_{x,\varepsilon}$ $|f_n(x) - f(x)| < \varepsilon$

(different order!)
different meaning

e.g. $X = [0,1]$

$f_n(x) = x^n$

The f_n converge pointwise to f, where $f(x) = \begin{cases} 0 & \text{if } x<1 \\ 1 & \text{if } x=1 \end{cases}$

But f_n does not converge uniformly to f

Pf of non-UC

If convergence were uniform,

I pick $\varepsilon = 0.01$ You pick $N = 10^6$

Is N big enough to make

$|f_n(x) - f(x)| < 0.01 \quad \forall x \in X$

$\forall n > 10^6$

I choose $x = (0.01)^{\frac{1}{10^6+1}} \sim 0.999999\ldots$

$n = 10^6 + 1$

$f(x) = 0$

$|f_n(x) - f(x)| = |x^n| = \left[(0.01)^{\frac{1}{10^6+1}}\right]^{10^6+1} = 0.01$

Remarks on uniform convergence

① If $f_1, f_2, \ldots$ converges uniformly to f, it converges uniformly to f

Pf Given UC, we know that $\forall \varepsilon > 0$, $\exists N_\varepsilon$ s.t. for all $x \in X$,

$\forall n \in N_\varepsilon$, $|f_n(x) - f(x)| < \varepsilon$

This shows e.g. that $\lim_{n\to\infty} f_n(0.7) = f(0.7)$

② If $f_1, f_2, \ldots$ are bounded functions in X, then $f_1, f_2, \ldots \to f$ uniformly iff $\lim_{i\to\infty} f_i = f$ if the metric space of bounded functions with sup norm

Pf Uniform convergence says,

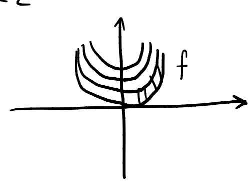
$\forall \varepsilon > 0$, $\exists N_\varepsilon$ s.t. $\forall n > N_\varepsilon$, $\underline{\forall x \in X, |f_n(x) - f(x)| < \varepsilon}$

$\Leftrightarrow \sup_{x\in X} |f_n(x) - f(x)| < \varepsilon$

$\Leftrightarrow \|f_n - f\|_{\sup} < \varepsilon$

But: consider a case like

$X = \mathbb{R} \quad f_n(x) = x^2 + \frac{1}{n}$



These converge uniformly to $f(x) = x^2$

But none of the f_n have a sup norm

But $f_n(x) = (1+\frac{1}{n})\, x^2$ would <u>not</u> converge uniformly to x^2

Uniform convergence of $f_n \to f$ equivalent to $f_n - f = 0$ in sup norm

③ $X = [0, c] \quad c < 1$

$f_n(x) = x^n$ <u>does</u> uniformly converge to 0

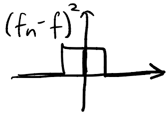
Given $\varepsilon > 0$, I need to find N_ε s.t.

$$\forall x \text{ in } [0,c] \quad |f_n(x) - f(x)| < \varepsilon$$
$$|f_n(x) - f(x)| = |x^n| < \varepsilon$$

It suffices to find N s.t. $|c^n| < \varepsilon$ for all $n > N$

take $N_\varepsilon = \dfrac{\log \varepsilon}{\log c} + 1000$

<u>Convergence in the L^2 sense</u>



$X = \mathbb{R}$

$$f_n(x) = \begin{cases} 1 & -\frac{1}{n} \le x \le \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

$f_n = \chi_{[-\frac{1}{n}, \frac{1}{n}]}$ ← characteristic function

What is the pointwise limit? $f = \chi_{\{0\}}$ $\quad f(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases}$

This convergence is <u>not</u> uniform: $f_n - f = \begin{cases} 1 & -\frac{1}{n} \le x < 0 \text{ or } 0 < x \le \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$

$$\|f_n - f\|_{\sup} = 1 \text{ for all } n$$

Now lets ask: is it the case that for all $\varepsilon > 0$, $\exists N_\varepsilon$ s.t. for all $n > N_\varepsilon$

$\|f_n - f\|_2 < \varepsilon$? $\qquad$ indeed, just take $N = 10000 \cdot \varepsilon^{-2}$

$$\|f_n - f\|_2 = \left(\int_{-\infty}^{\infty} [f_n(x) - f(x)]^2\, dx\right)^{1/2} = \left(\frac{2}{n}\right)^{1/2} = \frac{\sqrt{2}}{\sqrt{n}}$$

Your enemy argues: $f_1, f_2, \dots$ converges in L^2 to 0

$\|f_n - 0\|_2 = \|f_n\|_2 = \sqrt{\frac{2}{n}}$ which goes to 0 as $n \to \infty$

What is the distance between the two purported limits $\chi_{\{0\}}$ and 0?

$$\|\chi_{\{0\}} - 0\|_2 = \|\chi_{\{0\}}\|_2 = \left[\int_{-\infty}^{\infty} \chi_{\{0\}}^2(x)\, dx\right]^{1/2} = 0$$

Def We say a function $f: [a, b] \to \mathbb{R}$ is "L^2-null" if $\int_a^b |f(x)|^2\, dx = 0$

Def We say f and g are equivalent in L^2 if $f-g$ is L^2-null

$$f \sim_{L_2} g$$

Def The space of L^2 functions in $[a,b]$ is the set of convergence classes for $\sim_{L_2}$

Remarks

① Why is $\sim_{L_2}$ an equivalence class?

You need $f \sim g$ & $g \sim h \Rightarrow f \sim h$

$f \sim g$ & $g \sim h$ are L^2-null $\Rightarrow f \sim h$

really, we prove

if F, G are L^2-null, so is $F+G$

Indeed, the L^2-null functions are a subspace,

and L^2 space is the quotient of functions $/L^2$-null space

In fact, $f \sim g$