Size-Noise Tradeoffs in Generative Networks

Bolton Bailey   Matus Telgarsky

Univ. of Illinois Urbana-Champaign

November 29, 2018

Generative networks

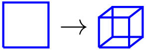
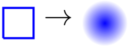
Easy distribution $X \in \mathbb{R}^n$.
Hard distribution $Y \in \mathbb{R}^d$.
Generator Network $g : X \to Y$.

What can Y be?

Previous Work

- Universal approximation theorem:
  Shallow networks approximate continuous functions.
- "On the ability of neural nets to express distributions":
  Upper bounds for representability & shallow depth separation.

Our Contribution: Wasserstein Error Bounds

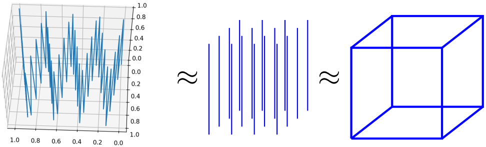
- $(n < d)$ Tight error bounds $\approx (\text{Width})^{\text{Depth}}$ 
  This is a *deep* lower bound.
- $(n = d)$ Switching distributions $\approx \text{polylog}(1/\text{Error})$. 
- $(n > d)$ Trivial networks approximate normal by addition.

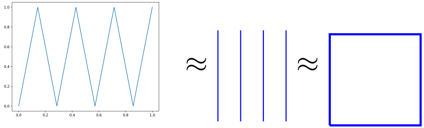
Increasing Uniform Noise ($n < d = kn$)

Networks going from Uniform $[0,1]^n$ to $[0,1]^{kn}$:

$$\text{Optimal Error} \approx (\text{Width})^{-\left(\frac{\text{Depth}}{k-1}\right)}.$$

Upper Bound Proof: Space filling curve.
Lower Bound Proof: Affine piece counting.

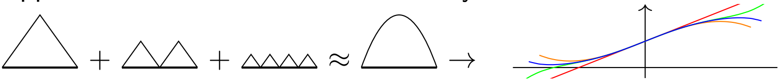




Normal ↔ Uniform ($n = d = 1$)

Normal → Uniform: Upper Bound

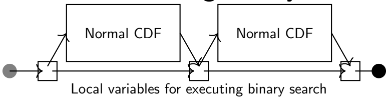
Approximate the normal CDF with Taylor series.



Size = polylog(1/Error).

Uniform → Normal: Upper Bound

Approximate the inverse CDF using binary search.



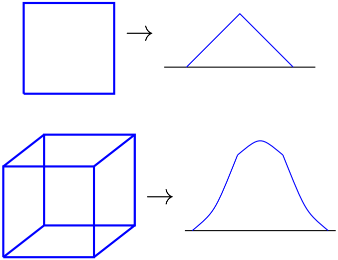
Size = polylog(1/Error).

Lower bounds

Size > log(1/Error) with more affine piece counting.

High Dimensional Uniform to Normal ($n > d$)

Summing independent uniform distributions approximates a normal.



With a version of Berry-Esseen, we have:

$$\text{Error} \approx 1/\sqrt{\text{Number of inputs.}}$$

Poster 10:45 AM - 12:45 PM Room 210 & 230 AB #141