Does Data Augmentation Lead to Positive Margin?

Zhili Feng
Zachary Charles
Po-Ling Loh
Dimitris Papailiopoulos

Shashank Rajput

* Equal Contribution
Data Augmentation (DA)

• DA means increasing the training set artificially.
• Used to train state of the art deep models.

Rotations, crops

Noise
Why use Data Augmentation (DA)?

Aim: 
*Build a model that is robust to slight perturbations of input*

Idea: 
*Train on perturbed versions of the inputs!*

*Works in practice! But can we prove it?*
Setup

- What margin does $w'$ achieve with respect to $S'$?
Setup

- **No DA**
  - Enforces no margin $\Rightarrow$ Not robust

- Blackbox learner – Outputs ANY classifier that fits the training set

- What margin does $\omega'$ achieve?
Setup

- $S$ → $S'$ → $w'$
  - DA → Augmented Dataset → Learning → Model
  - What margin does $w'$ achieve?

Blackbox learner – Outputs ANY classifier that fits the training set

No DA
- Enforces no margin ➔ Not robust

With DA
- Enforces some margin ➔ Robust
Can we use DA to enforce margin?
Can we use DA to enforce margin?

Idea: Create an $\varepsilon$-net of DA points.
Problem: $\varepsilon$-net requires exponentially many points
What is the minimum number of points we need?

Class 1

Class 2

Theorem: \(d+1\) points necessary and sufficient to get max-margin.
What is the minimum number of points we need?

Theorem: \( d+1 \) points necessary and sufficient to get max-margin.

Caveat: You need to know the max margin classifier — Beats the purpose!
Random DA: Points on the sphere

- What should the radius $\delta$ be?
- How many DA points?
Random DA: Points on the sphere

Max margin = $\nu^*$

$\delta = O(\nu^*)$

$0(2^d)$

#DA Points

Margin Achieved
Random DA: Points on the sphere

Max margin = $\nu^*$

$$\delta = \tilde{O}(\nu^* \sqrt{d})$$

$$\delta = O(\nu^*)$$

#DA Points

$O(\text{poly}(d))$  $O(2^d)$
Beyond Linear Classifiers

- Similar results for classifiers which “respect” local convex hulls of training points.
- Example: Nearest neighbor classifier.

Future Work:

More structured augmentation

- How much robustness do cropping, rotation etc. add?

Adaptive augmentation

- What margin does Adaptive Data Augmentation (Adversarial Training) achieve?
Thank you

• Poster #155
  • 6:30 – 9:00 PM, Today
  • Pacific Ballroom

• Emails: rajput3@wisc.edu, zfeng49@cs.wisc.edu