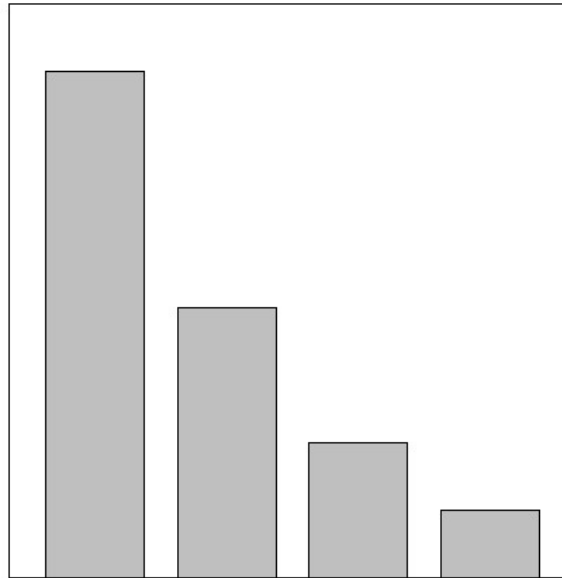


A systematic study of the class imbalance problem in CNN CNNにおけるクラス不均衡問題の体系的研究

- データのクラス分布が不均衡な場合、画像分類を行ってもデータの多いクラスにバイアスがかかり、うまく分類できない。



- 分類精度を向上させるためにはデータに対する対策とアルゴリズムに対する対策のどちらが効果的なのでしょうか？

A systematic study of the class imbalance problem in CNN CNNにおけるクラス不均衡問題の体系的研究

- MNIST、CIFAR-10、ImageNet (ILSVRC-2012) を利用して、不均衡な状態で実験を行ったところ、「**オーバーサンプリング**」が最も分類精度向上に効果があった。
- オーバーサンプリング：
少数クラスのデータをランダムにコピーすることで増やしてデータ数を多数クラスに合わせる。
⇔アンダーサンプリング：データ数を少ないクラスに合わせる。
- ただし、オーバーサンプリングはデータを単純に複製するだけなので、**過学習**に陥る可能性がある。
- 検証には、指標として不均衡なクラス分類で有効とされる**マルチクラスROC**を用いた。

参考文献

- Buda, Mateusz, Atsuto Maki, and Maciej A. Mazurowski. 2018. "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks." *Neural Networks: The Official Journal of the International Neural Network Society* 106 (October): 249–59.