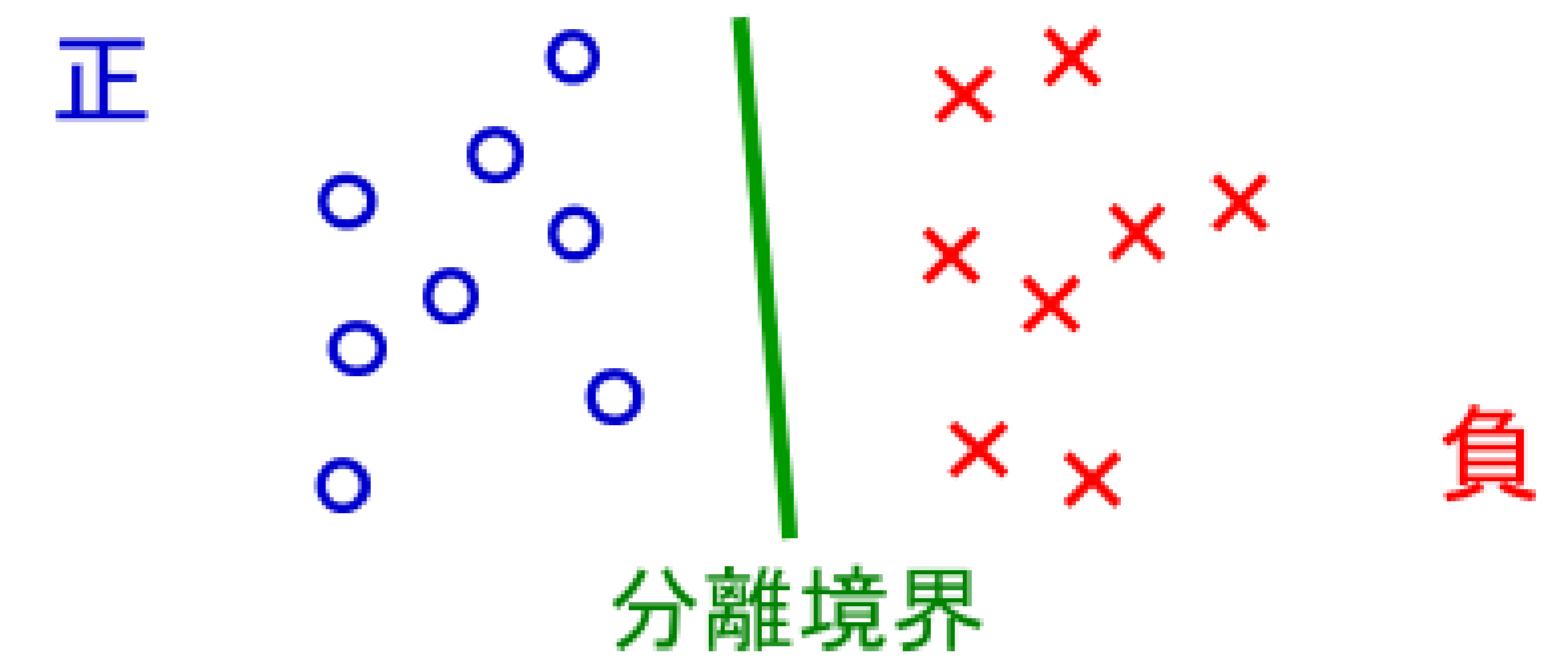


限られた情報からの機械学習

- 良質なビッグデータがあれば、現在の機械学習はヒトを超える性能が達成できる:
 - 画像理解, 音声認識, 言語翻訳, 商品推薦...
- しかし、応用分野によっては良質なビッグデータが簡単に取れない:
 - 医療データ解析
 - インフラの管理
 - 自然災害の防災・減災
 - 機能材料の開発
- そのような場面でも最先端の機械学習技術を活用したい!

2クラスの教師付き分類



- 大量のラベル付きデータを用いれば、精度良く分類境界が学習できる
- しかし、ラベル付きデータの収集には多大なコストがかかる
- コストの低い「弱い」データを活用する

正例とラベルなしデータからの分類

du Plessis, Niu & Sugiyama (NIPS2014, ICML2015)
 Niu, du Plessis, Sakai, Ma & Sugiyama (NIPS2016)
 Kiryo, du Plessis, Niu & Sugiyama (NIPS2017)

- 負例を集めるのが大変:
 - クリック vs. 非クリック
 - 友達 vs. 非友達
-
- ラベルなし(正と負の混合)
- 正例とラベルなしデータだけから、最適な分類ができる
 - 偏って集められた負例も活用できる
- Hsieh, Niu & Sugiyama (ICML2019)

教師なし分類

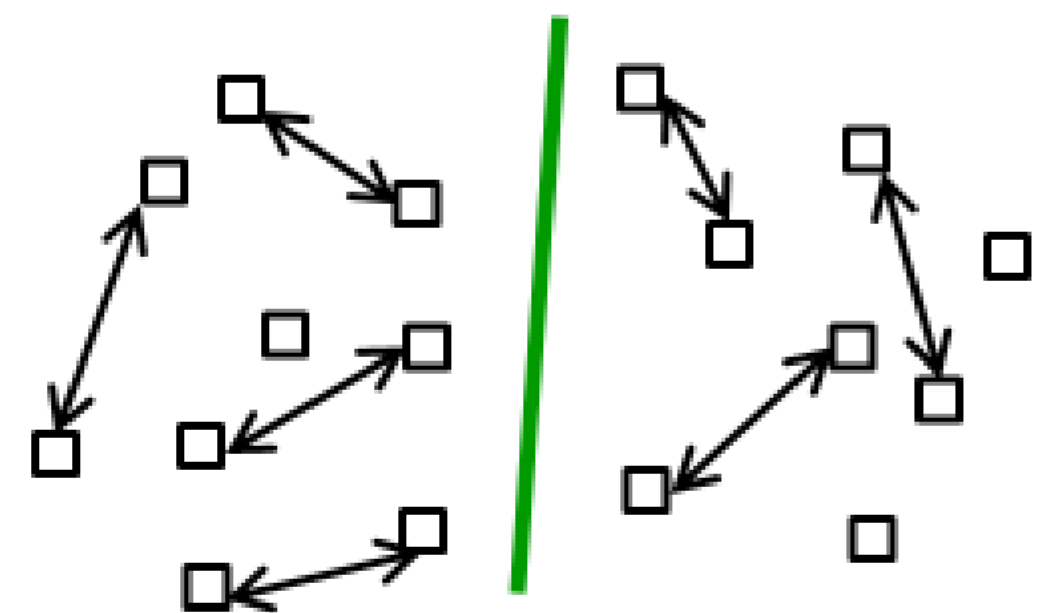
du Plessis, Niu & Sugiyama (TAAI2013)
 Lu, Niu, Menon & Sugiyama (ICLR2019)

- 完全に教師なしでも、**クラス比の異なるラベルなしデータが2セットあれば、最適な分類ができる**
-

類似データ対からの分類

Bao, Niu & Sugiyama (ICML2018)

- 財産, 宗教, 政治など, デリケートな質問に対して、明示的に趣向を回答するのははばかれる
 - 「あのと同じ」であれば回答しやすい!
- 類似データ対とラベルなしデータだけから、最適な分類ができる

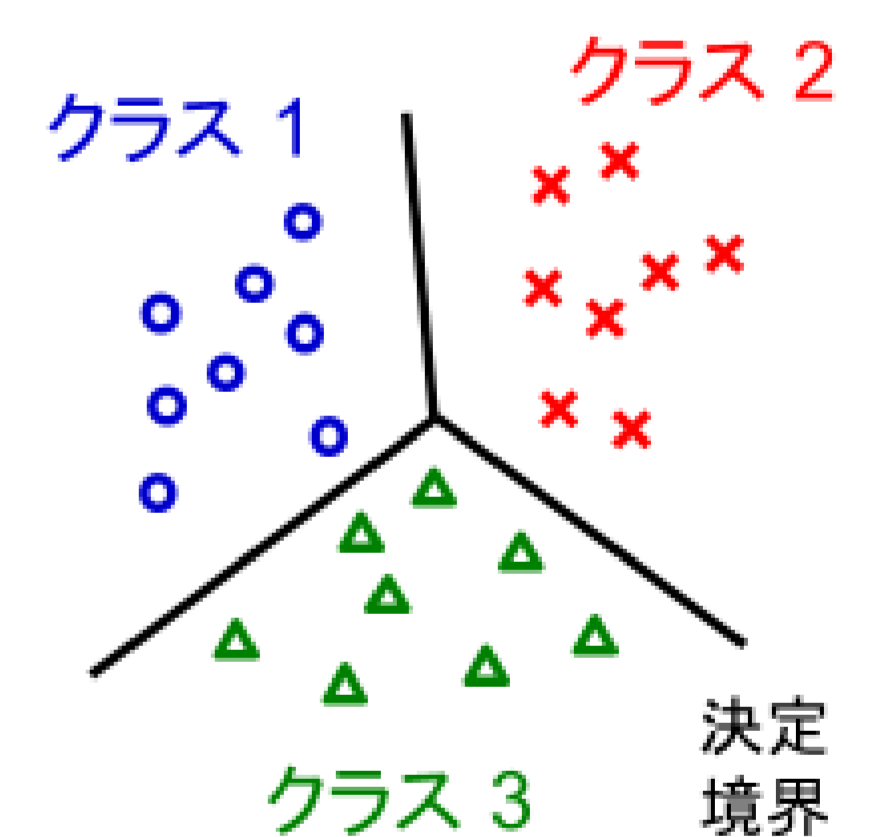


- 非類似データ対からでも分類できる
- Shimada, Bao, Sato & Sugiyama (arXiv2019)

補ラベルからの分類

Ishida, Niu & Sugiyama (NIPS2017)
 Ishida, Niu, Menon & Sugiyama (ICML2019)

- 多クラスの訓練データのラベル付けは高コスト
 - 多数の候補クラスから正しいものを選ぶ必要がある
- 補ラベル:
 - パターンが属さないクラスのラベル(補ラベル)を与える
 - 補ラベルをつけるのは低コスト
- 「間違ったラベル」だけから、最適な分類ができる!



チームメンバー

- 杉山将(チームリーダー)
- Gang Niu(研究員)
- Voot Tangkaratt(特別研究員)
- Miao Xu(特別研究員)
- Bo Han(特別研究員)



2019年度の主な研究成果

- Ni, C., Charoenphakdee, N., Honda, J., & Sugiyama, M. On the calibration of multiclass classification with rejection. *NeurIPS2019*.
- Xu, L., Honda, J., Niu, G., & Sugiyama, M. Uncoupled regression from pairwise comparison data. *NeurIPS2019*.
- Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., & Sugiyama, M. Are anchor points really indispensable in label-noise learning? *NeurIPS2019*.
- Charoenphakdee, N., Lee, J., Jin, Y., Wanvarie, D., & Sugiyama, M. Learning only from relevant keywords and unlabeled documents. *EMNLP-IJCNLP2019*.
- Charoenphakdee, N., Lee, J., & Sugiyama, M. On symmetric losses for learning from corrupted labels. *ICML2019*.
- Hsieh, Y.-G., Niu, G., & Sugiyama, M. Classification from positive, unlabeled and biased negative data. *ICML2019*.
- Ishida, T., Niu, G., Menon, A., & Sugiyama, M. Complementary-label learning for arbitrary losses and models. *ICML2019*.
- Wu, Y.-H., Charoenphakdee, N., Bao, H., Tangkaratt, V., & Sugiyama, M. Imitation learning from imperfect demonstration. *ICML2019*.
- Yu, X., Han, B., Yao, J., Niu, G., Tsang, I., & Sugiyama, M. How does disagreement help generalization against label corruption? *ICML2019*.
- Lu, N., Niu, G., Menon, A. K., & Sugiyama, M. On the minimal supervision for training any binary classifier from only unlabeled data. *ICLR2019*.
- Osa, T., Tangkaratt, V., & Sugiyama, M. Hierarchical reinforcement learning via advantage-weighted information maximization. *ICLR2019*.
- Charoenphakdee, N. & Sugiyama, M. Positive-unlabeled classification under class prior shift and asymmetric error. *SDM2019*.