

## Anomaly Detection vs. Supervised Learning

We started to use some label data with examples that we knew were either anomalous or not anomalous with  $y = 1$  or  $y = 0$ .

If we have labeled data:

- why do not use a supervised learning algorithm (logistic regression or neural networks)?
- when do we use anomaly detection and when do we use supervised learning?

### Use anomaly detection when...

- **We have a very small number of positive examples ( $y = 1$  ... 0 - 20 examples is common)**
  - Not enough data to "learn" positive examples
  - Save positive examples just for CV and test set
- **A large number of negative ( $y = 0$ ) examples (to fit  $p(x)$ ).**
  - Only need negative examples for this
- **We have many different "types" of anomalies and it is hard for any algorithm to learn from positive examples what the anomalies look like; future anomalies may look nothing like any of the anomalous examples we've seen so far.**
  - For Example: Many types of SPAM
    - For the spam problem, usually enough positive examples
    - So this is why we usually think of SPAM as supervised learning

Application and why they're anomaly detection:

- **Fraud detection**
  - If we have many different types of ways for people to try to commit fraud
  - and a relatively small number of fraudulent users on our website
  - then we use an anomaly detection algorithm
  - If we're a major on line retailer/very subject to attacks, sometimes might shift to supervised learning
- **Manufacturing (e.g. aircraft engines)**
  - If we make huge volumes maybe have enough positive data: use a supervised learning algorithm
- **Monitoring machines in a data center**
- ...

### Use supervised learning when...

- **We have a large number of both positive and negative examples. In other words, the training set is more evenly divided into classes.**
- **We have enough positive examples for the algorithm to get a sense of what new positives examples look like. The future positive examples are likely similar to the ones in the training set.**

Applications:

- Email spam classification
- Weather prediction (sunny/rainy/etc).
- Cancer classification
- ...

**Video Question:** Which of the following problems would you approach with an anomaly detection algorithm (rather than a supervised learning algorithm)? Check all that apply.

You run a power utility (supplying electricity to customers) and want to monitor your electric plants to see if any one of them might be behaving strangely.

- You run a power utility and want to predict tomorrow's expected demand for electricity (so that you can plan to ramp up an appropriate amount of generation capacity).

A computer vision / security application, where you examine video images to see if anyone in your company's parking lot is acting in an unusual way.

- A computer vision application, where you examine an image of a person entering your retail store to determine if the person is male or female.