Anomaly Detection and Recommender Systems

Sidharth Baskaran

June 2021

Anomaly detection problem

- Decide if a example is an anomaly to dataset
- Build model for $p(x)$ which is probability of feature being anomalous
  - $p(x) < \epsilon \implies$ anomaly and $p(x) \geq \epsilon$ is fine
  - Ex: fraud detection where $x^{(i)}$ is features of user i

Gaussian/Normal Distribution

- If $x \in \mathbb{R}$, then if x is a distributed Gaussian with mean μ and variance σ^2
  - $x \tilde{} \mathcal{R}(\mu, \sigma^2)$
  - Formula is $p(x; \mu, \sigma^2)$
  - Density formula is $p = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$
- Distribution is centered at μ and σ determines width
- Area under curve is always 1, so $\sigma \propto \text{height}^{-1}$
- Parameter estimation
  - $\mu = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$
  - $\sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu)^2$
  - Tend to use $\frac{1}{m}$ instead of $\frac{1}{m-1}$, both work equally well

Anomaly detection algorithm

- Model probability of each feature vector as $p(x) = p(x_1; \mu_1, \sigma_1^2) p(x_2; \mu_2, \sigma_2^2) \dots p(x_n; \mu_n, \sigma_n^2) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$
  - Assumes features are independent

$$p(x) = \prod_{j=1}^{n} p\left(x_j; \mu_j, \sigma_j^2\right) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{\left(x_j - \mu_j\right)^2}{2\sigma_j^2}\right)$$

- Is anomaly if $p(x) < \epsilon$

Developing and Evaluating an Anomaly Detection System

- Importance of real-number evaluation
  - Assume labeled data exists, anomalous and non-anomalous
  - Training set $x^{(i)}, \dots, x^{(m)} \rightarrow$ assume normal examples that are not anomalous
  - Define a cross validation and test set
- Fit model $p(x)$ on training set $\{x^{(i)}, \dots, x^{(m)}\}$
- On cross validation/test set example x, predict

$$y = \begin{cases} 1 \text{ if } p(x) < \epsilon \text{ (anomaly)} \\ 0 \text{ if } p(x) \geq \epsilon \text{ (normal)} \end{cases}$$

- Evaluation metrics
  - True positive, false positive, false negative, true negative
  - Precision/recall
  - F_1 score (if skewed)
  - Classification accuracy is not a good metric due to skewedness
- Can also use the CV set to choose ϵ

Anomaly Detection vs. Supervised Learning

- Anomaly Detection
  - Very small number of positive examples
  - Large number of negative examples
  - Many different types of anomalies $\rightarrow$ cannot discern what anomalies look like from small positive examples
- Supervised learning
  - Large number of positive and negative examples
  - Enough positive examples for algorithm to discern a positive example
    - Later positive examples are similar to those in training set