DATA 1030: Anomaly Detection of Falling People

Kyle Lam, Brown University

10/23/2024

https://github.com/plumol/data1030-final-project

Dataset: Falling People

- Anomaly detection: how can we accurately predict for a very small positive class in a very unbalanced dataset?

- Background: This dataset was created and used to develop strategies for safer "smart" environments.
  - Sourcing[1]: This data comes from sensor activity taken from a care-independent smart home environment for elderly people.
  - 4 sensors (ankles, belt, and chest area) collected position data for participants performing tasks.

- Problem: Classification!

[1] Vidulin, V., Lustrek, M., Kaluza, B., Piltaver, R., & Krivec, J. (2010). Localization Data for Person Activity [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C57G8X.

EDA

- Dataset: 20 users of training data, 5 users of test data (~130,000 data points x 5 columns)
  - XYZ coordinates, one-hot encoding of sensor part, anomaly target class

Complication	Falling People Dataset
Non-iid dataset?	Yes! Time series data
Large dataset?	Maybe? Semi-large and unbalanced (~130,000 data points + low positive class)
Missing values?	No.

Anomalies are shown by large shifts in XYZ positions
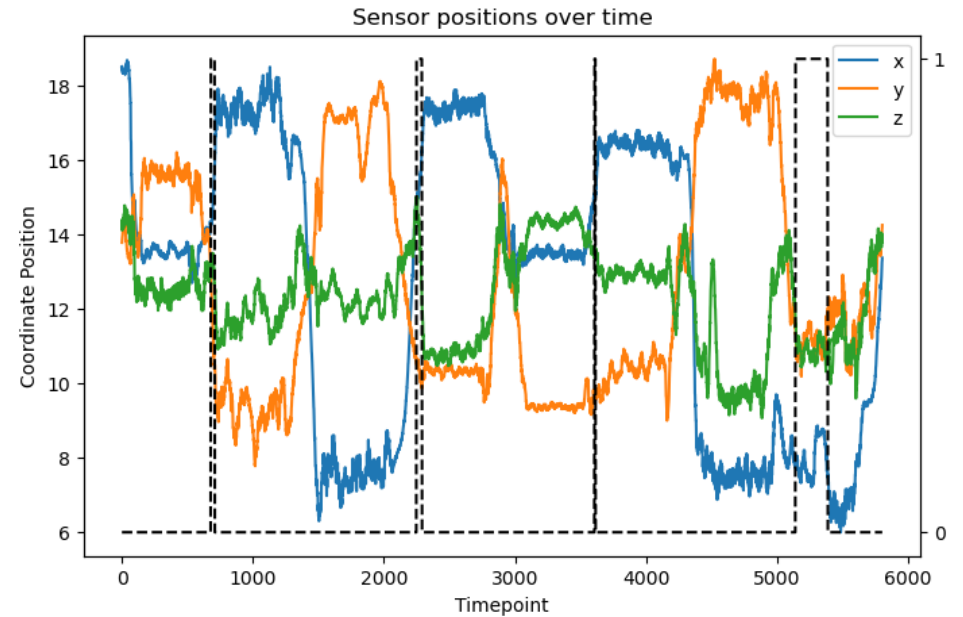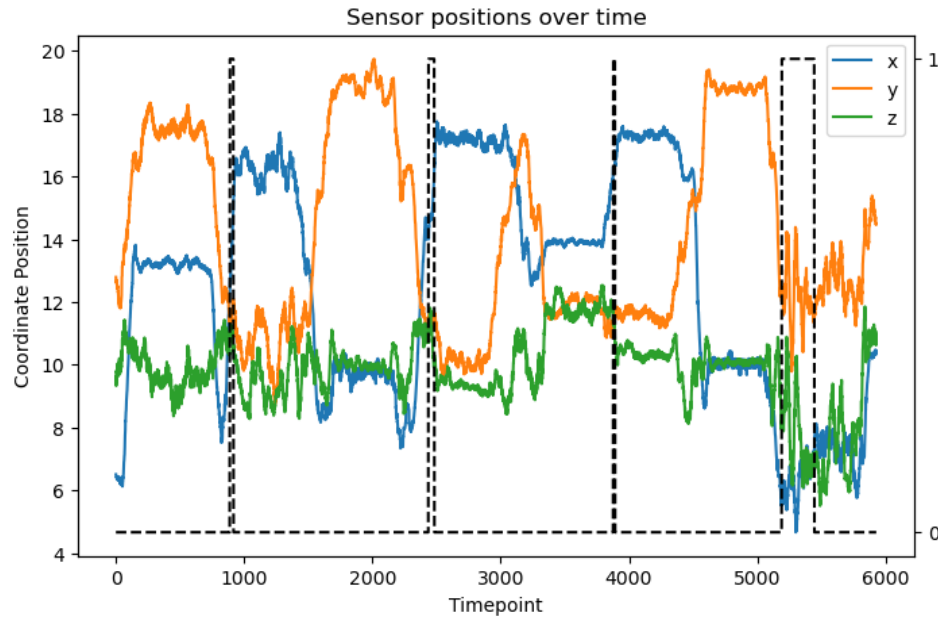
Different sensors show the same distribution

Sensor positions over time
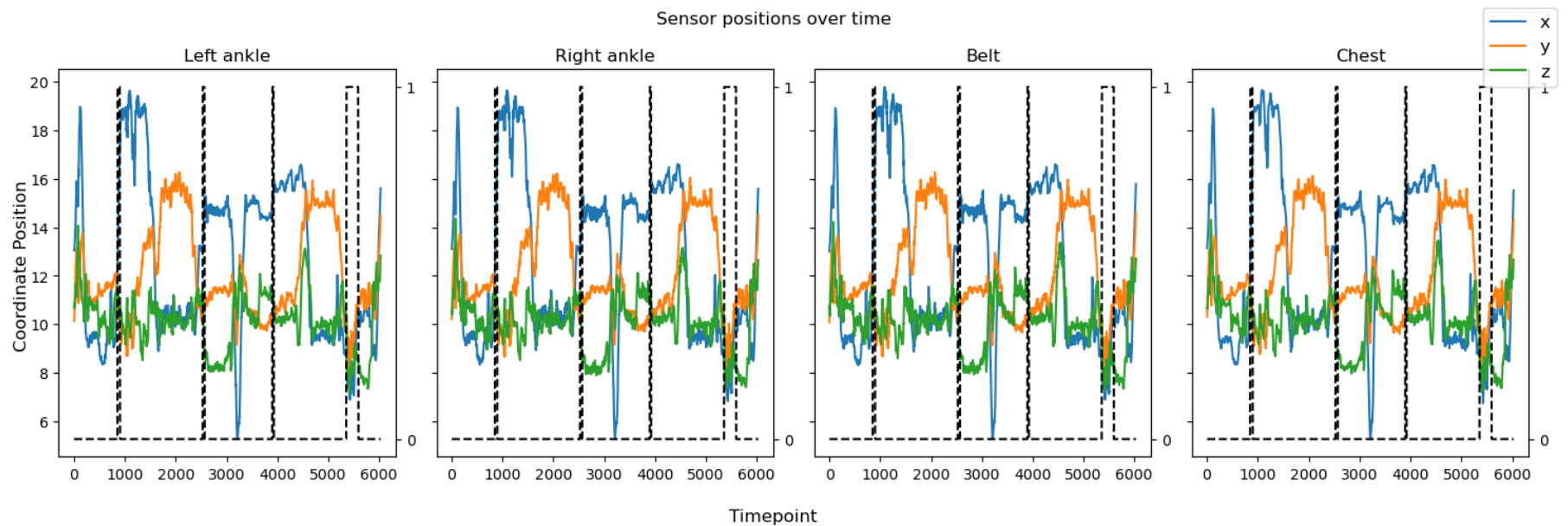
Different sensors show the same distribution

	x	y	z	010-000-024-033	010-000-030-096 \
3	13.110654	10.208509	10.591707	0.0	1.0
7	13.237308	10.412721	10.390196	0.0	1.0
10	13.324849	10.725250	10.453700	0.0	1.0
14	13.415006	11.155213	10.659015	0.0	1.0
19	13.574462	11.576481	10.992203	0.0	1.0
...	...	...	...	...	...
6019	14.914769	13.778870	11.770834	0.0	1.0
6022	15.136605	13.842691	12.317622	0.0	1.0
6025	15.359995	14.087712	12.329216	0.0	1.0
6028	15.548757	14.383471	12.485893	0.0	1.0
6031	15.586999	14.505781	12.649405	0.0	1.0

	x	y	z	010-000-024-033	010-000-030-096 \
0	13.045803	10.132120	10.657429	1.0	0.0
2	13.031613	10.161131	10.854713	1.0	0.0
6	13.176845	10.323796	10.673999	1.0	0.0
9	13.214779	10.582624	10.731702	1.0	0.0
13	13.336824	11.037615	10.909118	1.0	0.0
...	...	...	...	...	...
6004	13.954312	13.124557	11.885993	1.0	0.0
6010	14.331882	13.418463	11.769704	1.0	0.0
6017	14.788500	13.734064	11.974707	1.0	0.0
6023	15.252910	13.925743	12.106829	1.0	0.0
6030	15.612058	14.500909	12.832999	1.0	0.0

3D projections show anomaly points are clustered around a 'falling' event
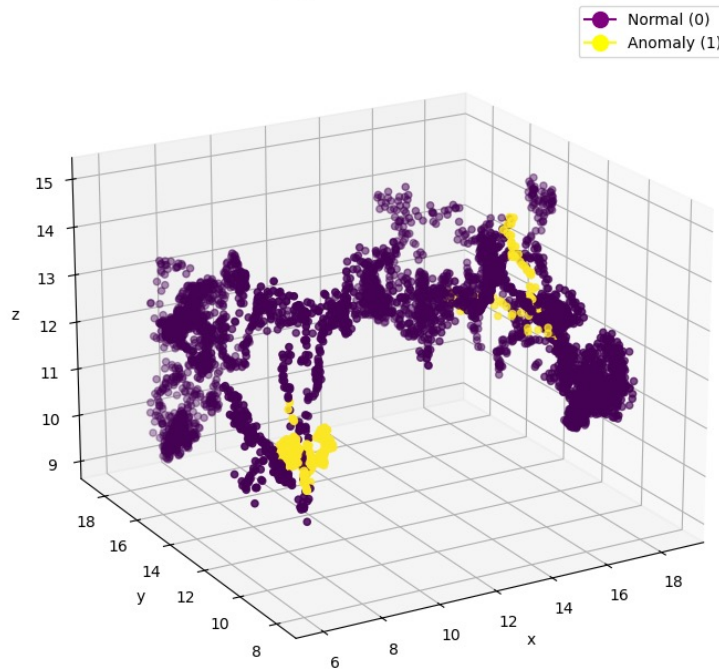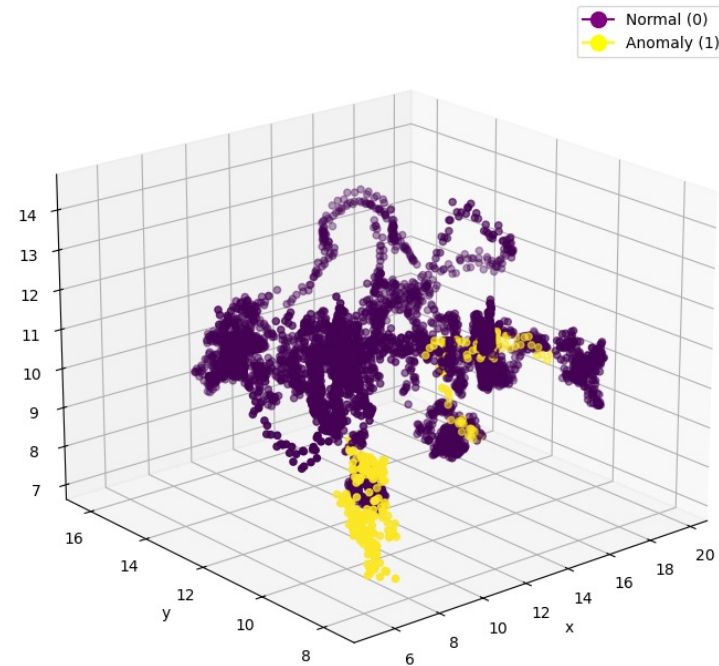
Preprocessing and Splitting

- This dataset is already split into training and testing

Type	Users	Samples	Normal %	Anomalous %
Training	20	134229 (127656 normal/ 6573 anomalous)	95.1%	4.9%
Testing	5	30030 (28420 normal/ 1610 anomalous)	94.6%	5.4%

- Dataset is already preprocessed and truncated
    - No missing values in original dataset
    - Timestamps were removed from original dataset
    - User files are concatenated to one large dataframe
- Normalization