

Ranking important sensor

1. **Problem:** This is a classification problem (two classes $y=1$ and $y=-1$). There are ten sensors (sensor0 to sensor9) which are features for classification.
2. **Algorithms:** We use popular classifiers: Support vector machine (SVM), Random forest (RF) and K-Neighbors (KNN) for this problem. We use Gridsearch to find the best parameter for each classifier. The method is to use ensemble method to get better result, i.e, we do all classifiers and average over them.

In order to rank all sensors, we do model selection both forward and backward, and then compare the results. When we use 3 classifiers with each sensor as only one feature dependence, we obtain a result

If we add some feature to feature “sensor6” we got a worse accuracy. That suggests us to do model selection in different way:

Forward selection: We add the least significant sensor in each iteration

Backward selection: We remove the most significant sensor in each iteration

3. **Result:** To get the consistent result, we iterate the algorithm m times and sum up the results (m could be 10, 20 and 30).
We get results (the number is sum up over ranking of sensors. Smaller is better ranking, bigger is worse ranking)

m=10

Ranking sensors with forward selection:

```
[('sensor6', 17), ('sensor8', 24), ('sensor4', 54), ('sensor2', 106),  
('sensor0', 109), ('sensor1', 152), ('sensor3', 175), ('sensor5', 204),  
('sensor9', 239), ('sensor7', 270)]
```

Ranking sensors with backward selection:

```
[('sensor6', 11), ('sensor8', 35), ('sensor4', 53), ('sensor2', 107),  
('sensor0', 118), ('sensor1', 156), ('sensor3', 172), ('sensor5', 202),  
('sensor9', 232), ('sensor7', 264)]
```

m=20

Ranking sensors with forward selection:

```
[('sensor6', 28), ('sensor8', 56), ('sensor4', 108), ('sensor2', 197),  
('sensor0', 224), ('sensor1', 318), ('sensor3', 346), ('sensor5', 411),  
('sensor9', 472), ('sensor7', 540)]
```

Ranking sensors with backward selection:

```
[('sensor6', 23), ('sensor8', 70), ('sensor4', 111), ('sensor2', 212),  
('sensor0', 237), ('sensor1', 308), ('sensor3', 337), ('sensor5', 400),  
('sensor9', 470), ('sensor7', 532)]
```

m=30:

Ranking sensors with forward selection:

```
[('sensor6', 46), ('sensor8', 78), ('sensor4', 173), ('sensor2', 303),  
('sensor0', 330), ('sensor1', 459), ('sensor3', 524), ('sensor5', 618),  
('sensor9', 709), ('sensor7', 810)]
```

Ranking sensors with backward selection:

```
[('sensor6', 29), ('sensor8', 106), ('sensor4', 161), ('sensor2', 325),  
('sensor0', 351), ('sensor1', 467), ('sensor3', 497), ('sensor5', 593),  
('sensor9', 718), ('sensor7', 803)]
```

We see that our algorithm converges to one ranking both for forward and backward selection algorithms.