Importance Of Feature Selection in ML Models or: How I Learned to Start Worrying and Love Ensembles

Sankalp Gilda
Graduate Student, Dept. of Astronomy
University of Florida
Feature Selection...

1) **What** is it?
2) **Why** do we need it?
3) **How** do we do it?
4) Does it **help**?
Feature Selection...

1) **What** is it?
2) **Why** do we need it?
3) **How** do we do it?
4) Does it **help**?
Feature Selection...

1) What is it?
2) Why do we need it?
3) How do we do it?
4) Does it help?
Because this.

Source: www.xkcd.com
And this.

I don't trust linear regressions when it's harder to guess the direction of the correlation from the scatter plot than to find new constellations on it.

Source: www.xkcd.com
DATA TSUNAMI IS APPROACHING

WHERE ARE MY FLOATIES?
Feature Selection...

1) **What** is it?

2) **Why** do we need it?

3) **How** do we do it?

4) Does it **help**?
Feature Selection...

1) **What** is it?
2) **Why** do we need it?
3) **How** do we do it? How did I do it?
4) Does it **help**?
<table>
<thead>
<tr>
<th>Predictors/Features</th>
<th>Target</th>
</tr>
</thead>
</table>

**Header**
Feature Selection...

1) **What** is it?
2) **Why** do we need it?
3) **How** do we do it? How did I do it?
4) Does it **help**?
Correlation Matrix for Best feature set, [Fe/H]
Feature Selection...

1) **What** is it?

2) **Why** do we need it?

3) **How** do we do it? How did I do it?

4) **Does it help?** ☑
Feature Selection...

1) What is it?

2) Why do we need it?

3) How do we do it? How did I do it?

4) Does it help? ✓

5) Is it robust?
Robustness

- Stress Testing
- Reproducibility
Feature Selection...

1) **What** is it?
2) **Why** do we need it?
3) **How** do we do it? How did I do it?
4) **Does it** help? ✓
5) **Is it** robust? ✓
What’s Next?

- Move to a Probabilistic Framework (80%)
- Clean up code, and free up hard-coded params (80%)
- Create User Guide (50%)
- Assess performance against resolution, SNR, #samples
- Post to GitHub by end of Nov.
Questions?

Please feel free to get in touch if you have any questions, concerns or suggestions!

email: s.gilda@ufl.edu