

Visualization for Machine Learning models

Alexey Zaytsev,
Skoltech, CDISE
16 January

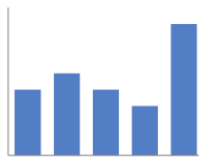
There is a lot of basic plots

91%

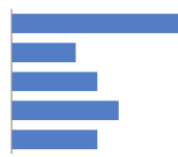
Simple text



Scatterplot



Vertical bar



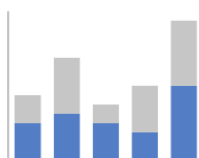
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

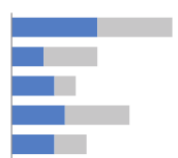
Table



Line



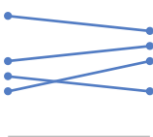
Stacked vertical bar



Stacked horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

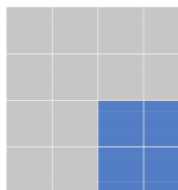
Heatmap



Slopegraph



Waterfall



Square area

FIGURE 2.1 The visuals I use most

Knafllic, C.N. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015

We need specific plot for particular application areas

**They are just basic plots, but with
specific curves or points displayed**

We consider machine learning as an example

Main problem statements in Machine Learning

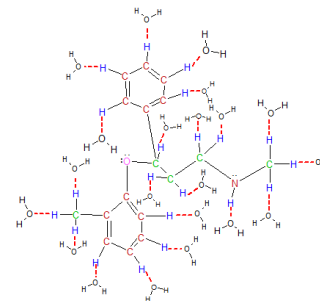
- Regression
- Classification
- Clustering
- Dimension reduction

For regression target is continuous

We want to predict the target continuous value

Examples:

- Predict solubility of a molecule
- Predict price of a house
- Predict performance of a portfolio

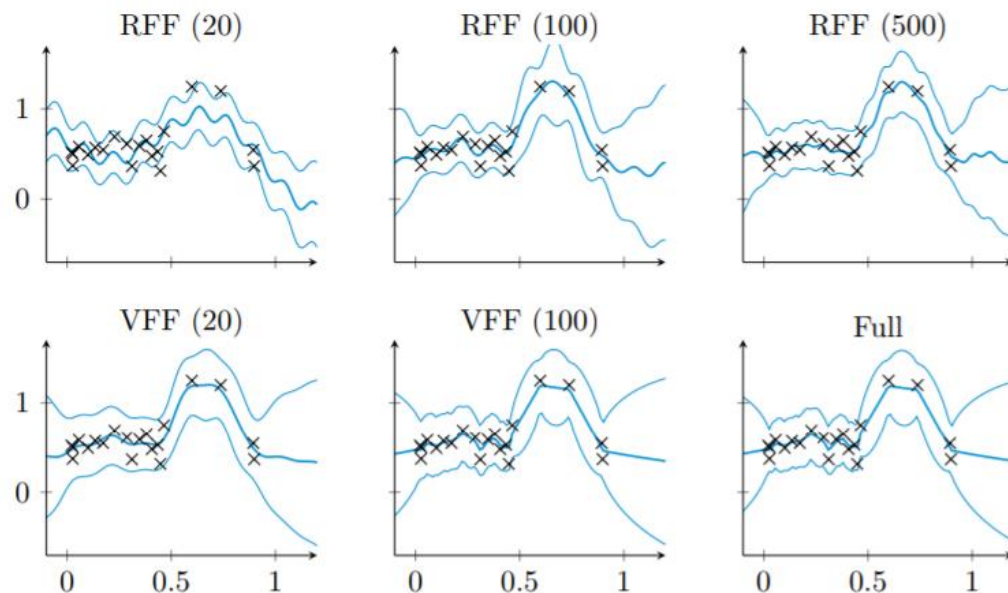


Simplest way: measure errors for different approach and get table

N	10,000		100,000		1,000,000		5,929,413	
	MSE	NLPD	MSE	NLPD	MSE	NLPD	MSE	NLPD
VFF	0.89 ± 0.15	1.362 ± 0.091	0.82 ± 0.05	1.319 ± 0.030	0.83 ± 0.01	1.326 ± 0.008	0.827 ± 0.004	1.324 ± 0.003
Full-RBF	0.89 ± 0.16	1.349 ± 0.098	N/A	N/A	N/A	N/A	N/A	N/A
Full-additive	0.89 ± 0.16	1.362 ± 0.096	N/A	N/A	N/A	N/A	N/A	N/A
SVIGP	0.89 ± 0.16	1.354 ± 0.096	0.79 ± 0.05	1.299 ± 0.033	0.79 ± 0.01	1.301 ± 0.009	0.791 ± 0.005	1.300 ± 0.003
String GP [†]	1.03 ± 0.10	N/A	0.93 ± 0.03	N/A	0.93 ± 0.01	N/A	0.90 ± 0.01	N/A
rBCM [†]	1.06 ± 0.10	N/A	1.04 ± 0.04	N/A	N/A	N/A	N/A	N/A

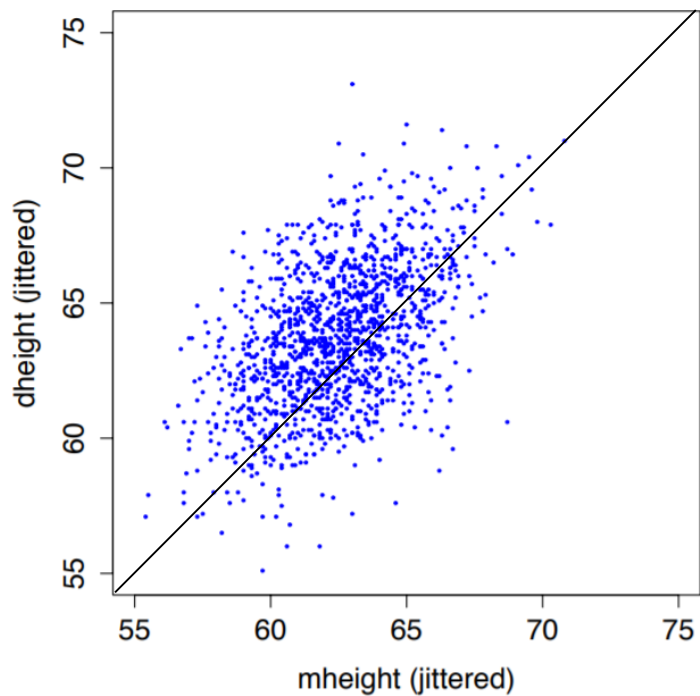
Hensman, James, Nicolas Durrande, and Arno Solin. "Variational Fourier features for Gaussian processes." *Journal of Machine Learning Research* 18, no. 151 (2018): 1-52.

Another simple way is to plot the model itself



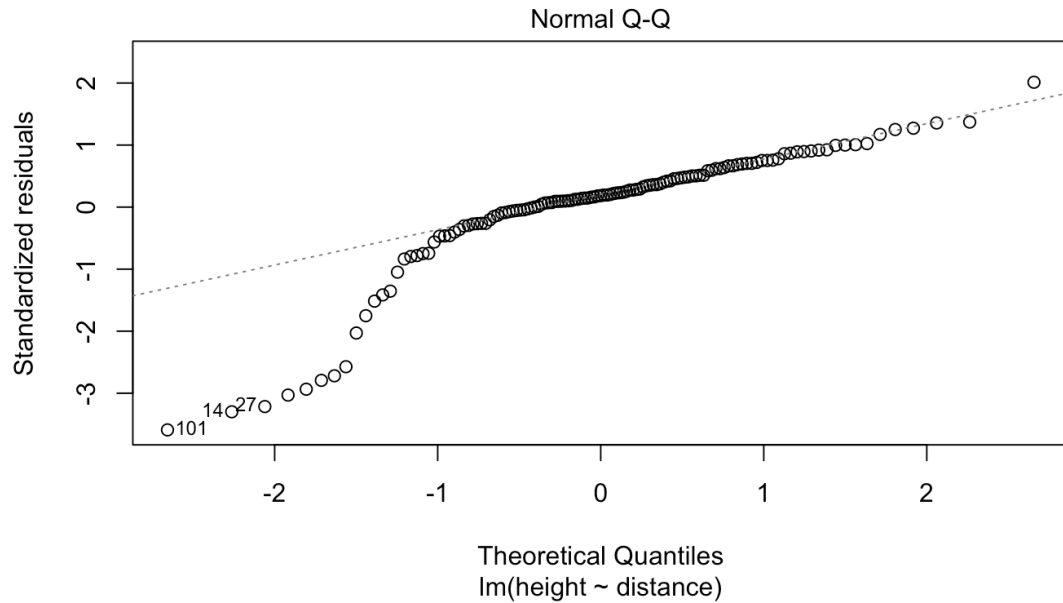
Hensman, James, Nicolas Durrande, and Arno Solin. "Variational Fourier features for Gaussian processes." *Journal of Machine Learning Research* 18, no. 151 (2018): 1-52.

Scatter plot



<http://staff.washington.edu/dbp/s423/PDFs/01-chapter-ALR-for-printing.pdf>

QQ plot



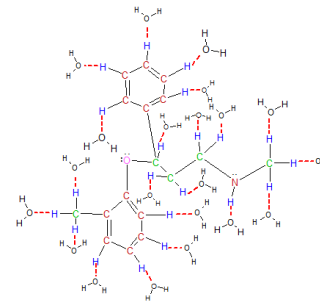
<https://stats.stackexchange.com/questions/253916/what-does-this-q-q-plot-indicate-about-my-data>

For classification target is discrete

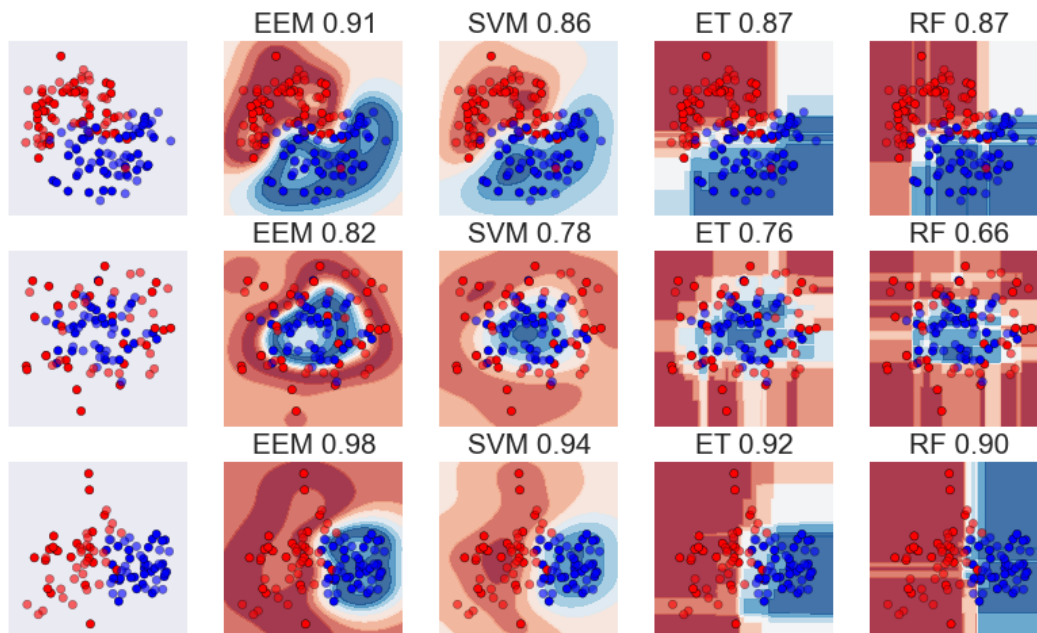
We want to predict the target class of an object

Examples:

- Is the molecule a good drug?
- Will the customer pay a mortgage?
- What dog breed is at the figure?

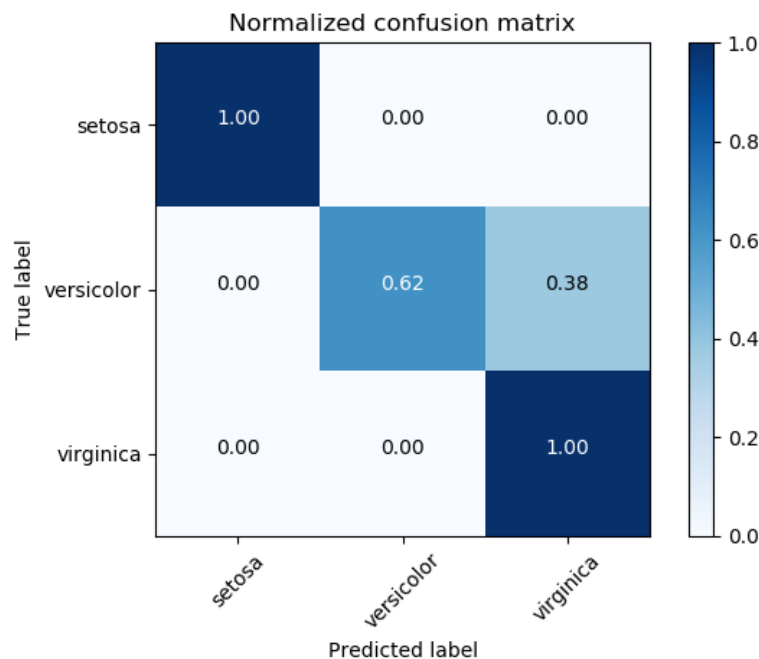


If data is 2D we can plot the model

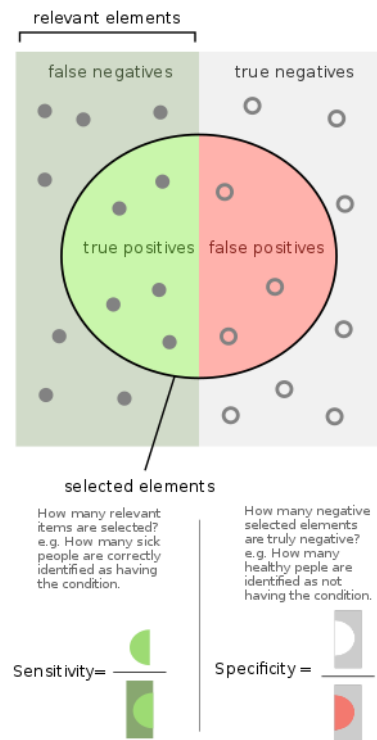


https://www.researchgate.net/publication/283740661_Extremely_Randomized_Machine_Learning_Methods_for_Compound_Activity_Prediction

Simplest way: calculate errors

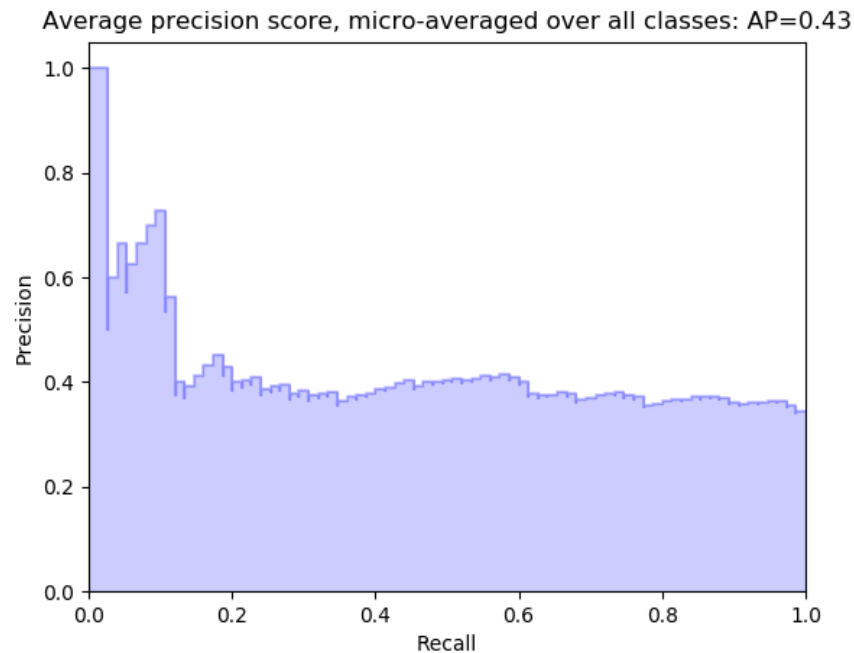
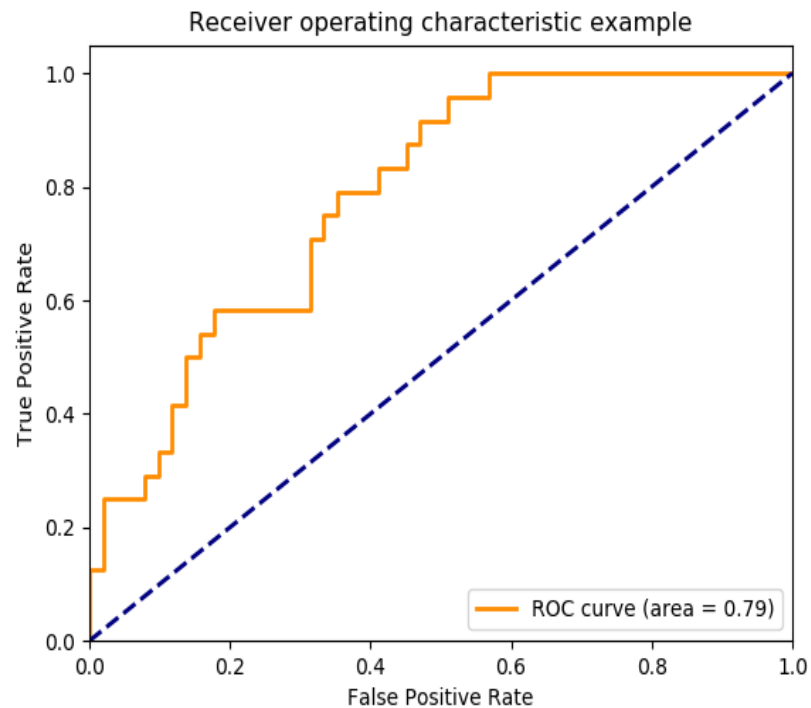


https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

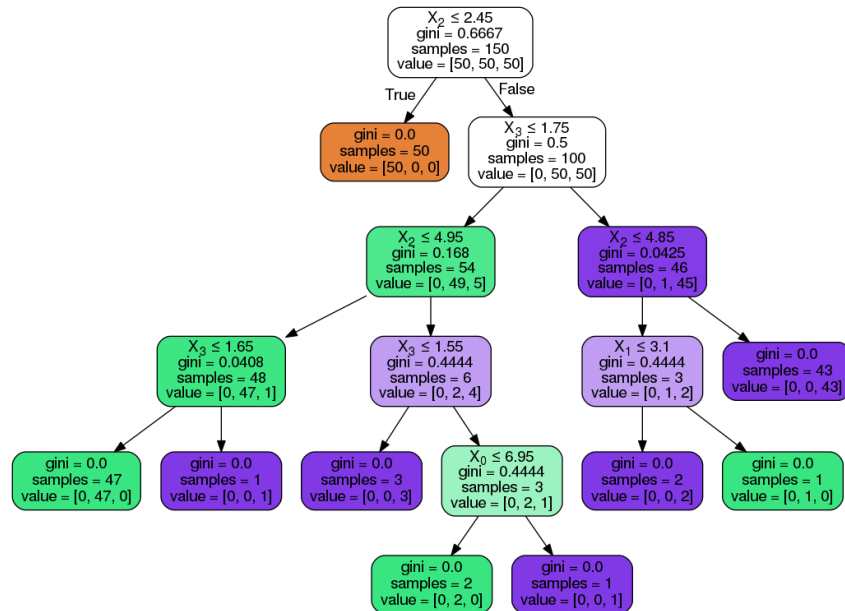


https://en.wikipedia.org/wiki/File:Sensitivity_and_specificity.svg

Simplest way: ROC & Precision Recall curves



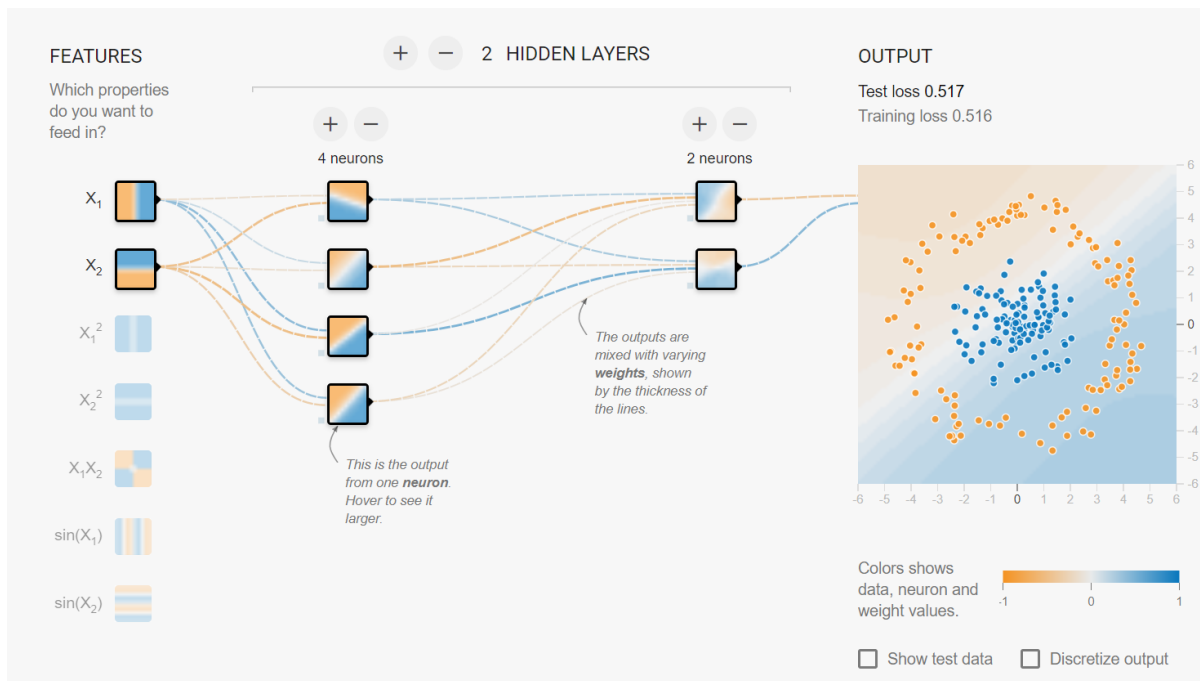
Visualize the model: decision tree



<https://medium.com/@rnbrown/creating-and-visualizing-decision-trees-with-python-f8e8fa394176>

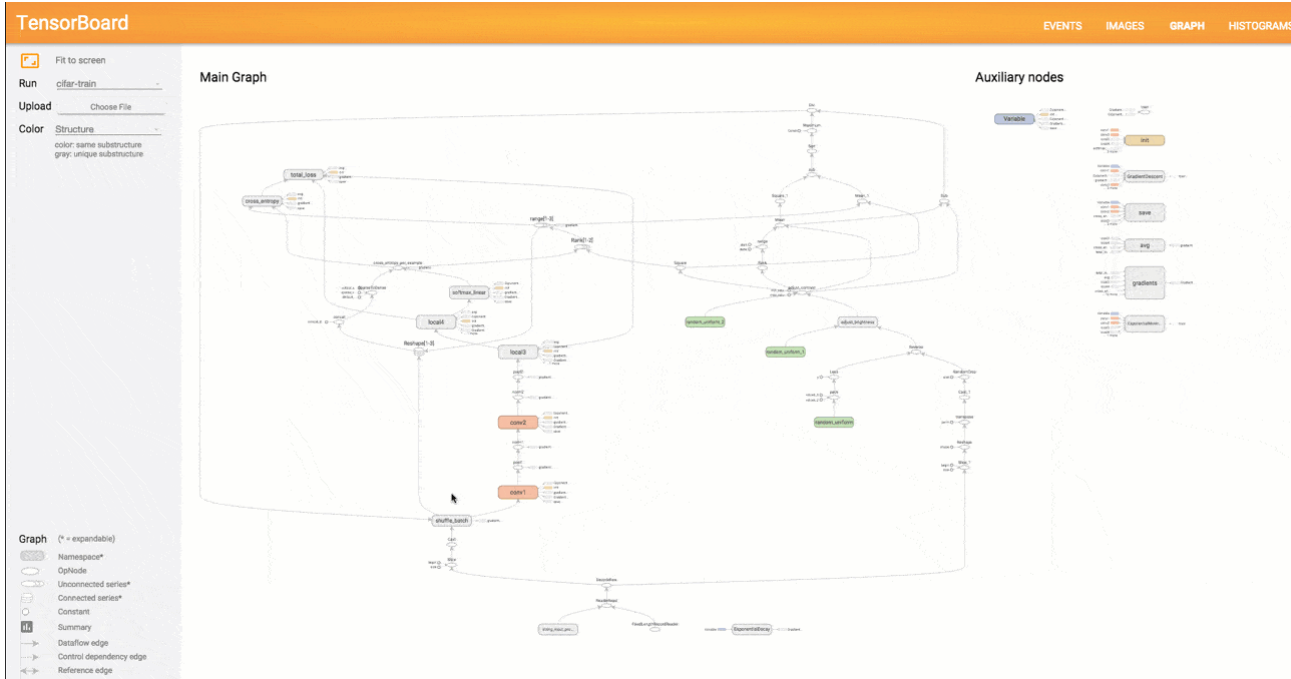


Visualize the model: Neural Network



<https://playground.tensorflow.org/>

Visualize the model: Neural Network with Tensorboard



https://www.tensorflow.org/guide/graph_viz

Linear regression

```
print(results.summary())
```

```
OLS Regression Results
```

Dep. Variable:	y	R-squared:	0.075
Model:	OLS	Adj. R-squared:	0.026
Method:	Least Squares	F-statistic:	1.532
Date:	Sat, 02 Jul 2016	Prob (F-statistic):	0.216
Time:	15:16:59	Log-Likelihood:	-725.57
No. Observations:	61	AIC:	1459.
Df Residuals:	57	BIC:	1468.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	-7990.8552	1.18e+04	-0.679	0.500	-3.16e+04 1.56e+04
x1	6704.9269	6546.017	1.024	0.310	-6403.245 1.98e+04
x2	549.6695	458.838	1.198	0.236	-369.138 1468.477
x3	1.819e+04	1.48e+04	1.227	0.225	-1.15e+04 4.79e+04

Omnibus:	103.868	Durbin-Watson:	2.058
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2650.260
Skew:	5.181	Prob(JB):	0.00
Kurtosis:	33.584	Cond. No.	63.4

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

<http://www.michaeljgrogan.com/statsmodels-sklearn-linear-regression/ols-regression-results/>

Summary

- If you are into Machine learning there are specific plots
 - For regression it is scatter plot or box plot for cross-validation runs
 - For classification there are ROC and PR AUC curves
- There a lot of ways to visualize your machine learning models
 - Decision trees
 - Neural networks
 - Linear regression
 - ...