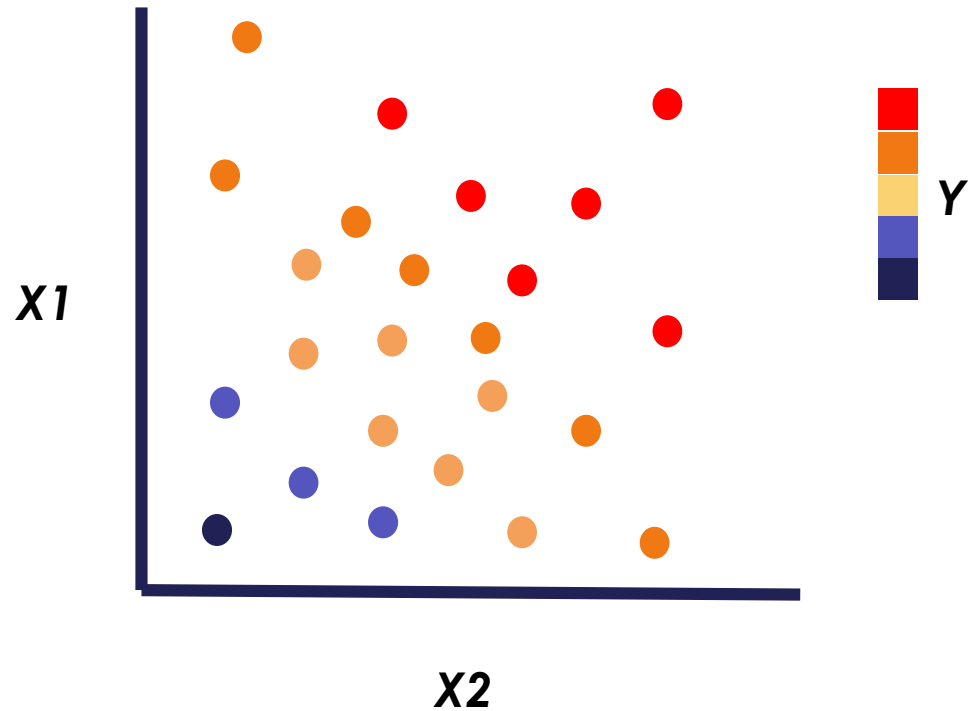


Accumulated local effects

Motivation

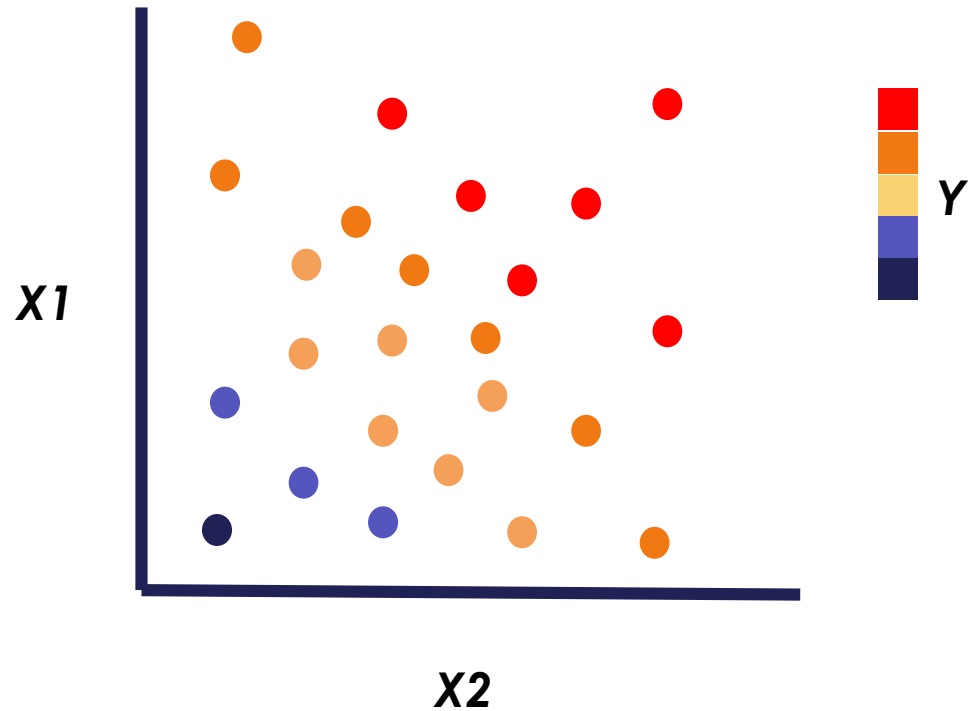


No correlation



Aim: Understand how a feature influences the prediction of a machine learning model on average.

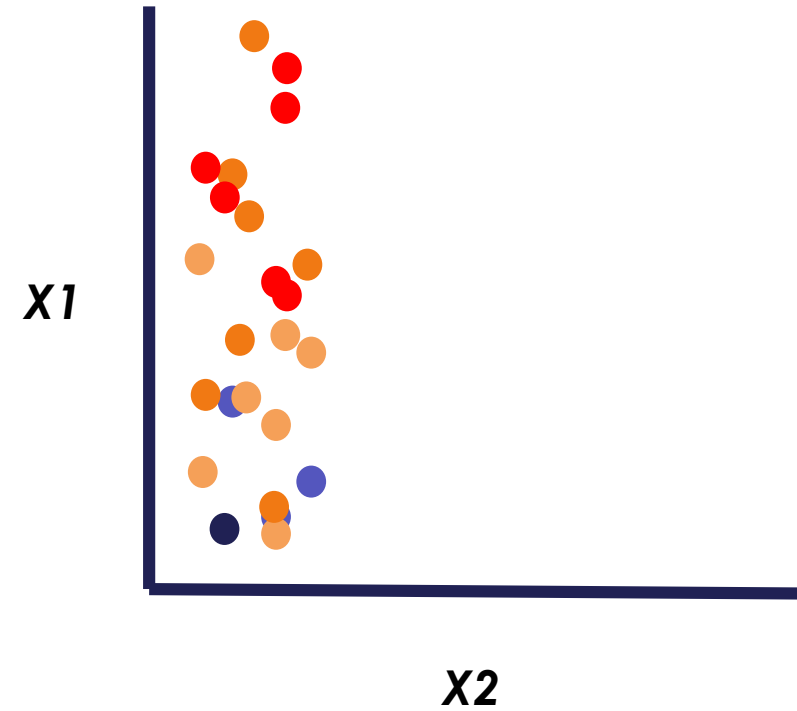
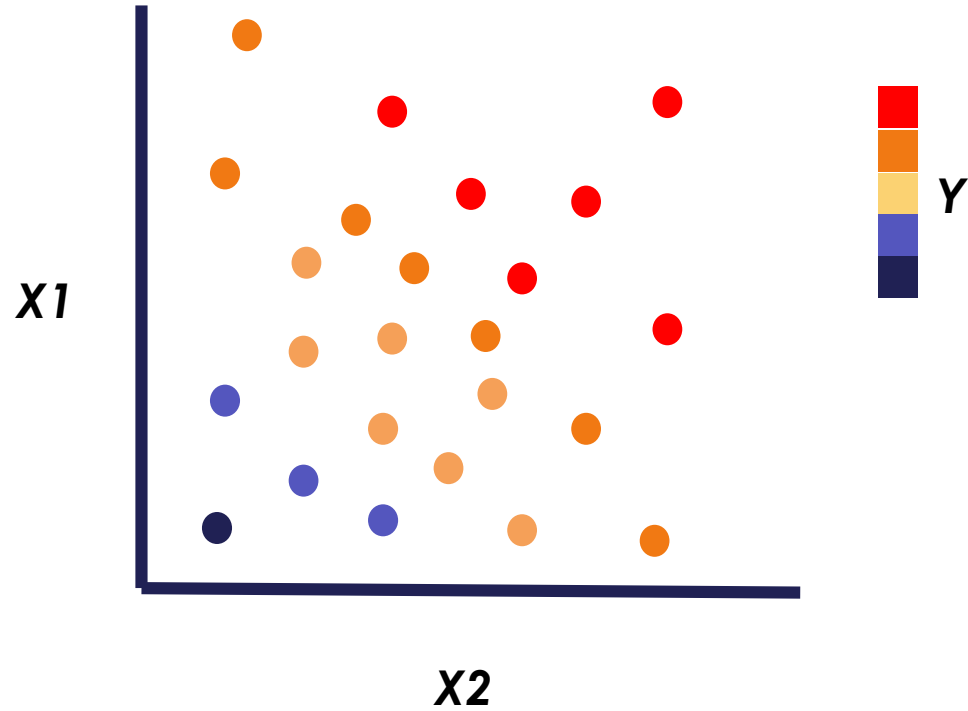
No correlation - PDP



PDP: Average prediction for various feature values.

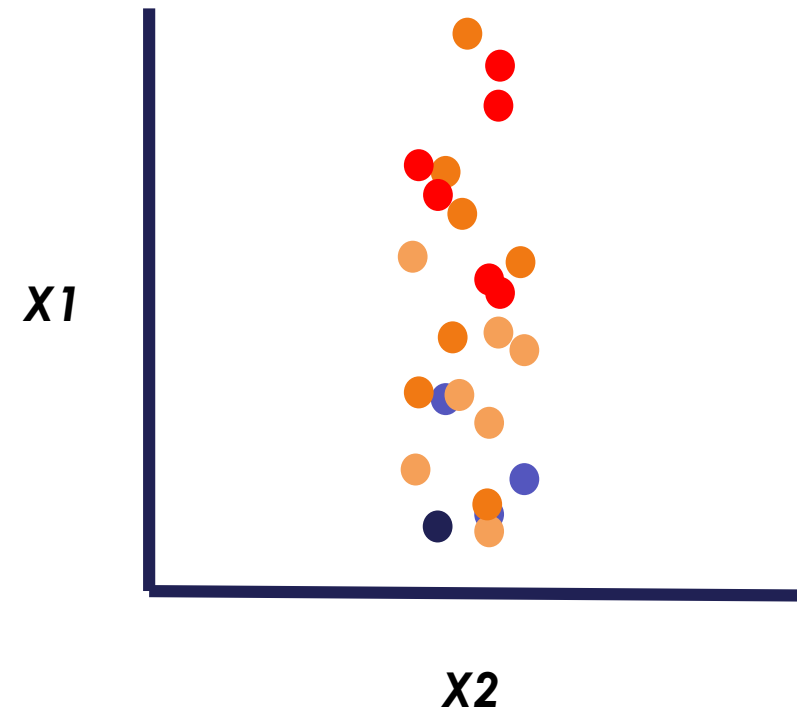
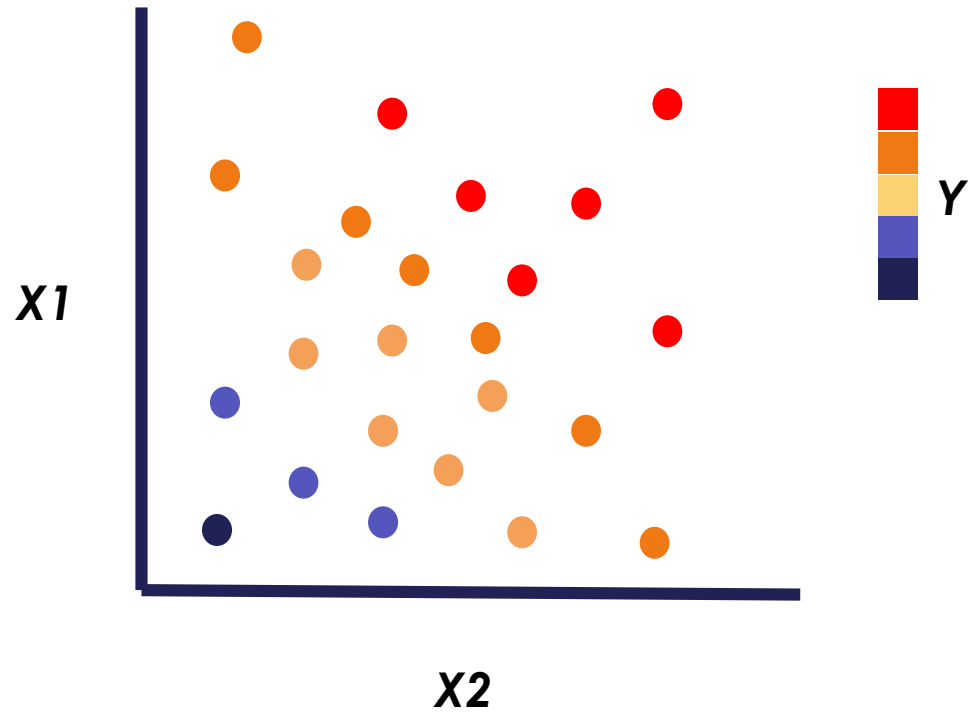
Set the feature to a single value, make predictions, take the average.

No correlation - PDP



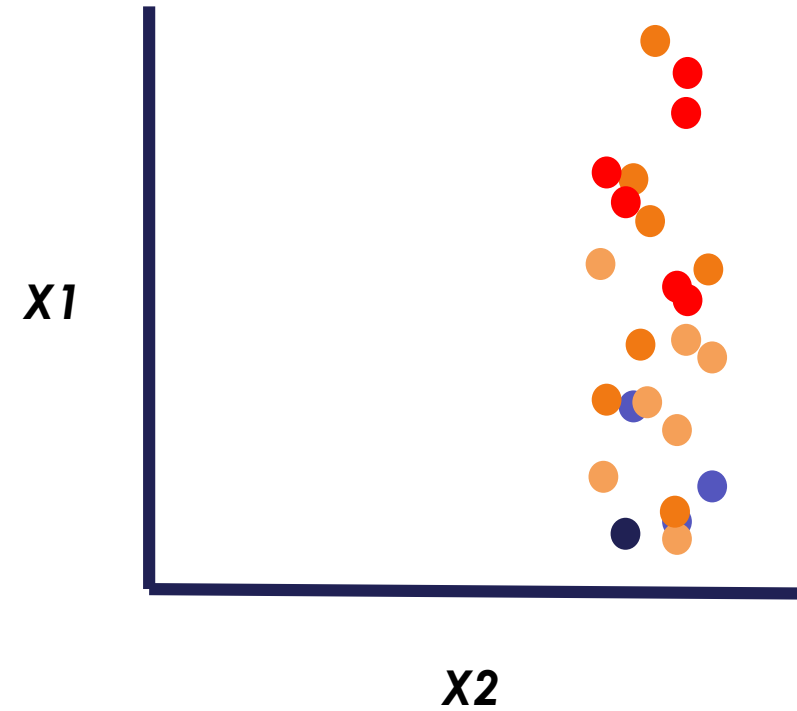
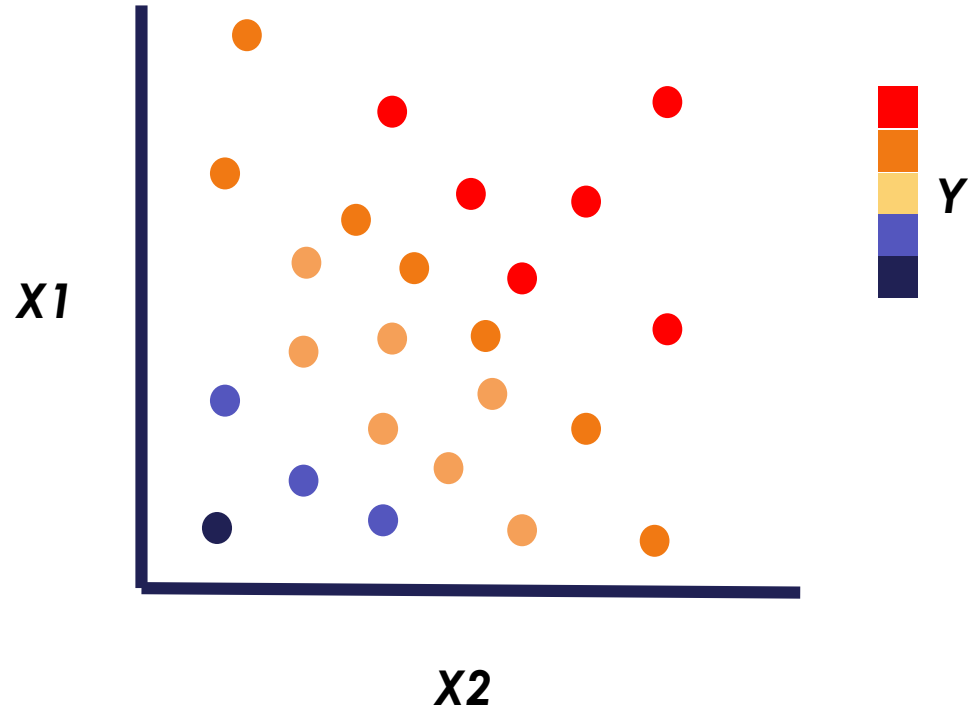
Set the X1 to a small value.

No correlation - PDP



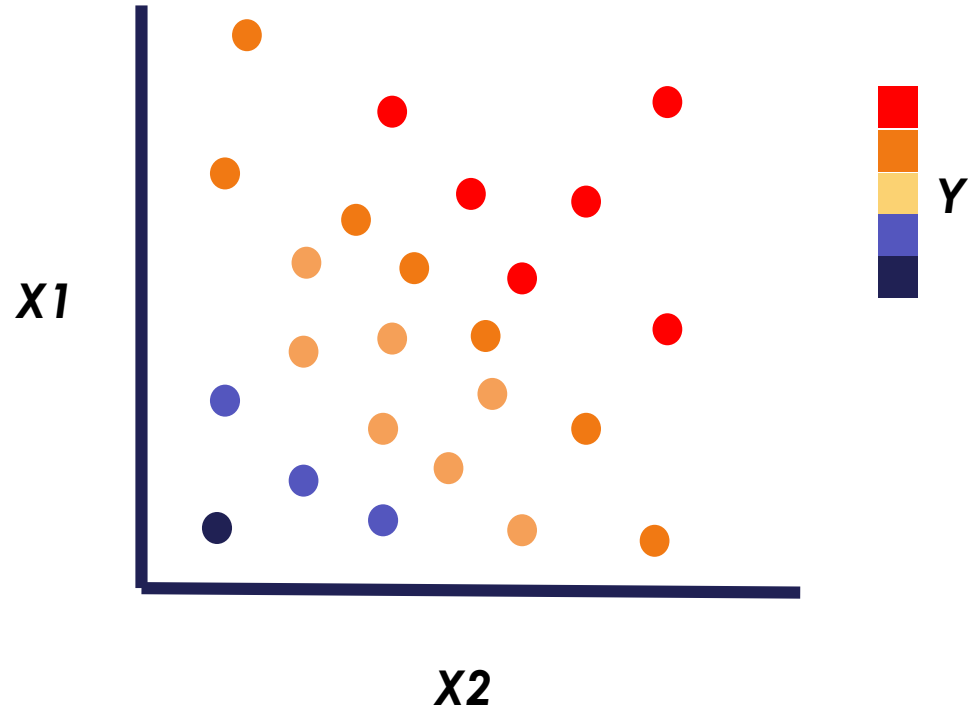
Set the X1 to a medium value.

No correlation - PDP



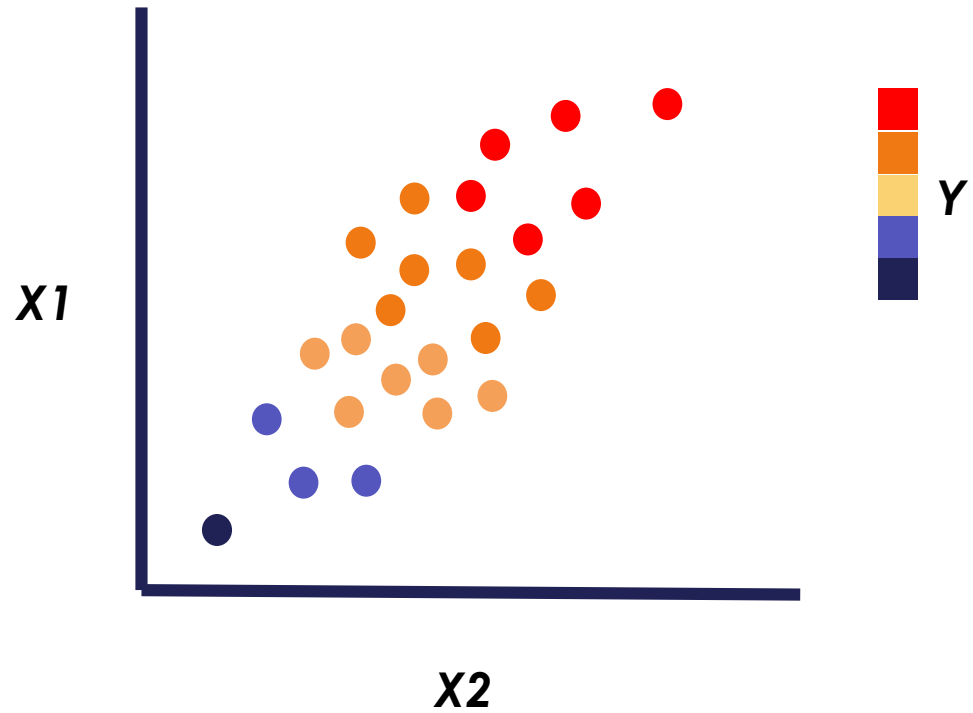
Set the X1 to a large value.

No correlation - PDP

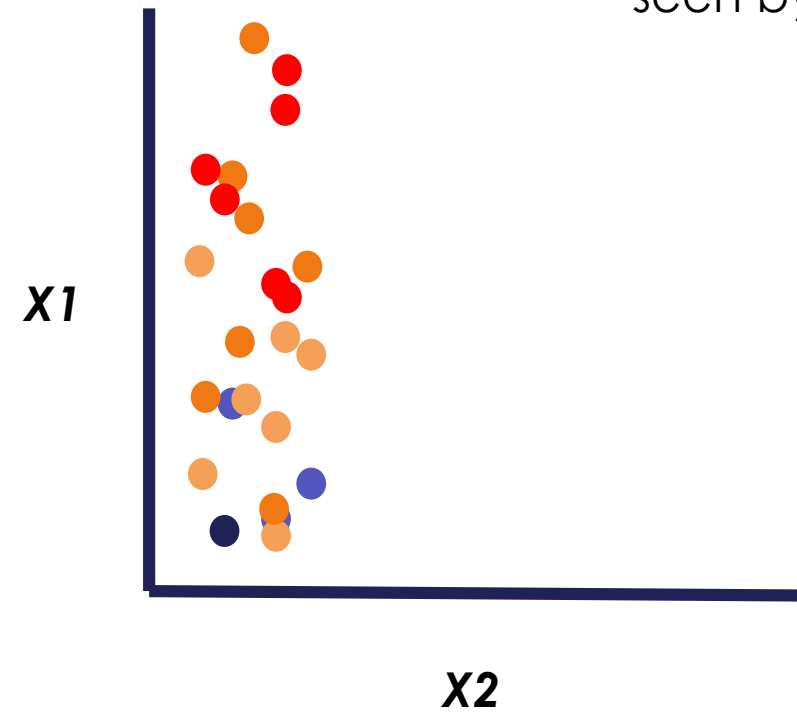
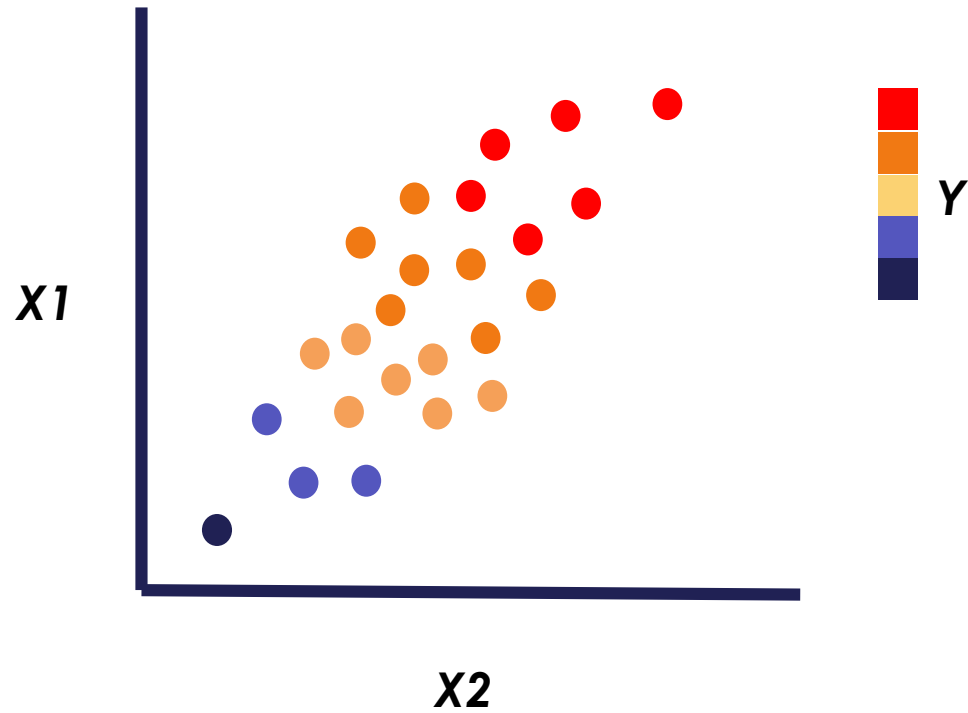


PDP: When features are not correlated, the model does not need to extrapolate to situations it has not seen already.

Correlation

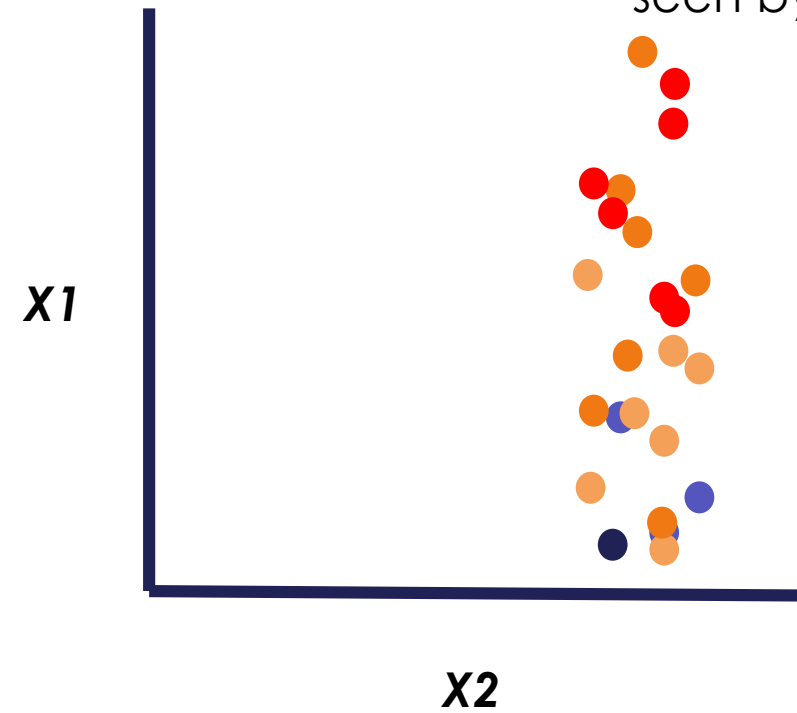
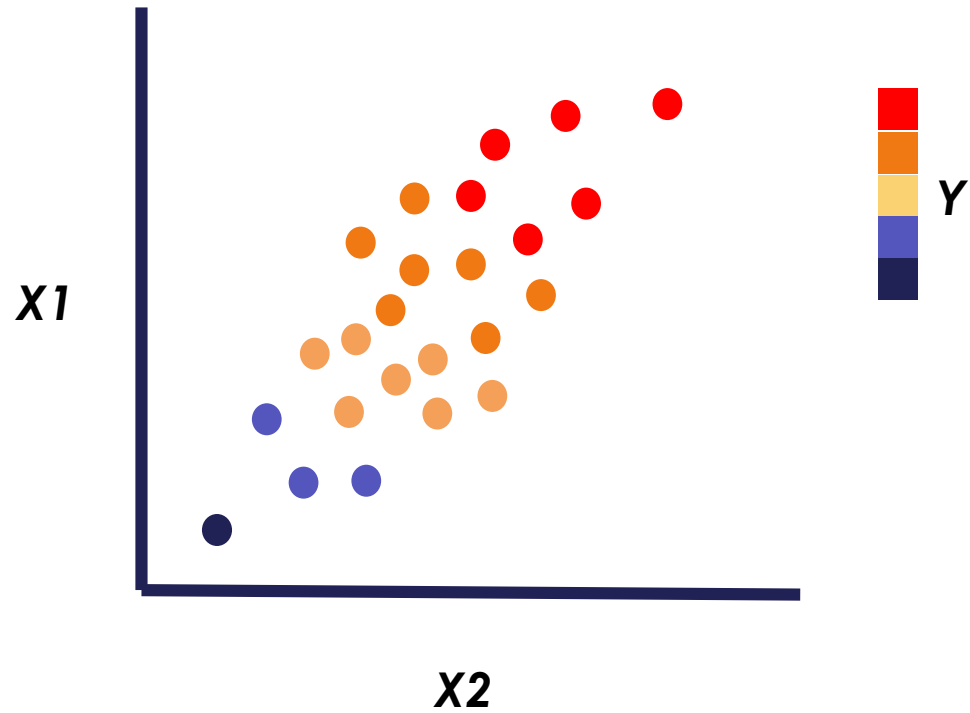


Correlation



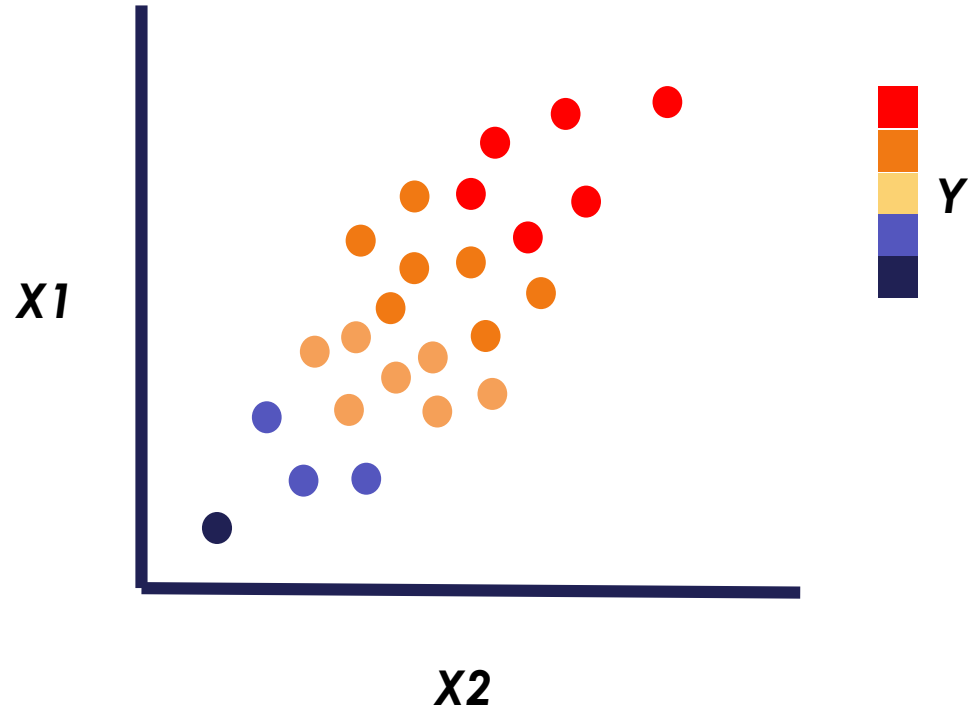
Setting X1 to certain values creates combinations not seen by the model.

Correlation



Setting X1 to certain values creates combinations not seen by the model.

Correlation



When features are correlated PDP cannot be trusted.

The model needs to “extrapolate” and make predictions for situations not encountered during training.

Tree based models are particularly bad at extrapolating.



Accumulated Local effects

Accumulated local effects plots were designed to be robust to correlated features, and resolve the problem of PDP plots.

THANK YOU

www.trainindata.com