Ethics of Data Science – Part II

Understanding interactions

What is an interaction, really?

- Consider a clinical trial with 20 patients, half of which are given a treatment T, and the other half are given placebo, in a randomized fashion.
- We are measuring an outcome where higher is worse (e.g., a pain or fatigue score).
- Then assume that only younger patients benefit from that treatment, with older adults (say, over 60) having no benefit from the treatment.

What is an interaction, really?

- Consider a clinical trial with 20 patients, half of which are given a treatment T, and the other half are given placebo, in a randomized fashion.
- We are measuring an outcome where higher is worse (e.g., a pain or fatigue score).
- Then assume that only younger patients benefit from that treatment, with older adults (say, over 60) having no benefit from the treatment.

```
          Younger    Older
Placebo   1.4006415 1.418481
Treatment 0.5429926 1.666739
```

What is an interaction, really?

- Consider a clinical trial with 20 patients, half of which are given a treatment T, and the other half are given placebo, in a randomized fashion.
- We are measuring an outcome where higher is worse (e.g., a pain or fatigue score).
- Then assume that only younger patients benefit from that treatment, with older adults (say, over 60) having no benefit from the treatment.

```
           Younger    Older
Placebo   1.4006415 1.418481
Treatment 0.5429926 1.666739
> summary(lm(y~treatment+age+treatment:age, data=Dint))

Call:
lm(formula = y ~ treatment + age + treatment:age, data = Dint)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5197 -0.1195  0.0129  0.1509  0.4427

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.40064    0.11146  12.567 1.05e-09 ***
treatment     -0.85765    0.15762  -5.441 5.44e-05 ***
age            0.01784    0.15762   0.113 0.911300
treatment:age  1.10591    0.22292   4.961 0.000142 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What is an interaction, really?

- Now assume that younger patients do better than older patients in general, and that the drug has a milder beneficial effect across all patients, regardless of age.

Drug only works for young patients

```
           Younger    Older
Placebo   1.4006415 1.418481
Treatment 0.5429926 1.666739
> summary(lm(y~treatment+age+treatment:age, data=Dint))

Call:
lm(formula = y ~ treatment + age + treatment:age, data = Dint)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5197 -0.1195  0.0129  0.1509  0.4427

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.40064    0.11146  12.567 1.05e-09 ***
treatment     -0.85765    0.15762  -5.441 5.44e-05 ***
age            0.01784    0.15762   0.113 0.911300
treatment:age  1.10591    0.22292   4.961 0.000142 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Drug works for everyone. Younger patients do better even without treatment.

```
           Younger    Older
Placebo   0.9006415 1.418481
Treatment 0.5429926 1.166739
> summary(lm(y~treatment+age+treatment:age, data=Dlin))

Call:
lm(formula = y ~ treatment + age + treatment:age, data = Dlin)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5197 -0.1195  0.0129  0.1509  0.4427

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.9006     0.1115   8.081 4.87e-07 ***
treatment     -0.3576     0.1576  -2.269  0.03746 *
age            0.5178     0.1576   3.285  0.00466 **
treatment:age  0.1059     0.2229   0.475  0.64113
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The presence of an interaction can matter a lot

*"A **prognostic** biomarker provides information about the patients overall cancer outcome, regardless of therapy"*

*"A **predictive** biomarker gives information about the effect of a therapeutic intervention. A predictive biomarker can be a target for therapy."*

Mistakenly treating a prognostic factor as predictive might lead to refusing treatment to patients that would benefit from it.

Conversely, mistakenly treating a predictive factor as prognostic, might lead to patients receiving treatment that does not have any benefit for them.

Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer. 2008 May;44(7):946-53. doi: 10.1016/j.ejca.2008.03.006. Epub 2008 Apr 7. PMID: 18396036.

What is an interaction, really?

An interaction in the binary case is literally a product:

$$x \cdot z = \begin{cases} 1, \text{ if and only if } x = z = 1, \\ 0, \text{ otherwise.} \end{cases}$$

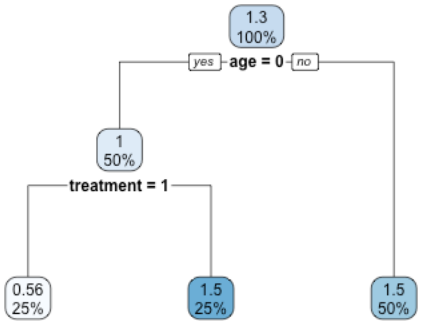
In the general case, an interaction is any non-additive function of two covariates, x and z:

$$y = f(x, z) \neq g_1(x) + g_2(z), \text{ for any functions } g_1, g_2$$

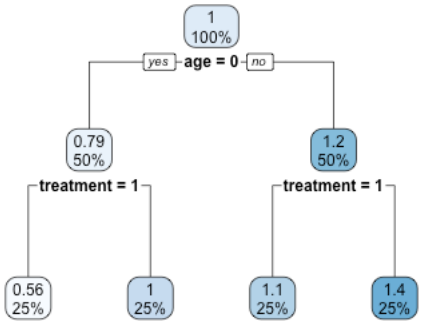
The above are 2-way interactions. A k-way interaction involves k variables instead. For example, a treatment that only works for younger adults with type II diabetes would manifest in a 3-way interaction term between age, the type of diabetes, and the treatment variable.

What is an interaction, really?

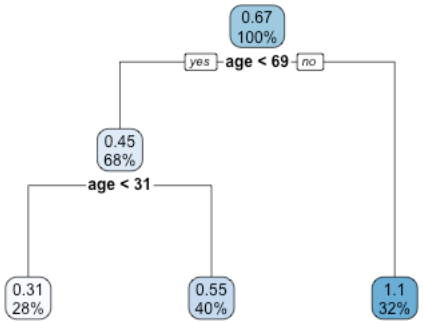
In general, a decision tree with maximum depth k will include k-way interaction terms. Decision trees are naturally explainable when it comes to interactions, but less so for additive effects.



Decision tree for dataset with interaction effect (treatment only works for older adults)



Decision tree for dataset with no interaction effect but additive age and treatment effects (younger patients do better both on or off treatment, treatment helps everyone)



Decision tree trained on age only, which in this case has a simple linear effect.

How can we detect an interaction in ML models?

We already discussed how differing shapes of ICE plots as well as vertical dispersion in SHAP dependence plots indicate interactions.

Another tool is SHAP interactions. These generalize SHAP values to measure the contribution of each pair of variables instead:

$$\sum_j \phi_j = f(\xi) - \frac{1}{n}\sum_{i=1}^{n} f(\vec{x}_i)$$ SHAP values

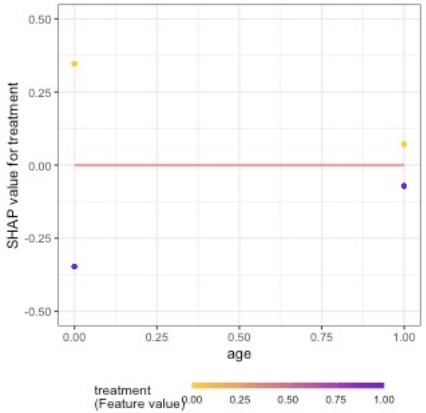
$$\sum_j \tilde{\phi}_j(\xi) + \sum_{j\neq k} \tilde{\phi}_{j,k}(\xi) = f(\xi) - \frac{1}{n}\sum_{i=1}^{n} f(\vec{x}_i)$$ SHAP interactions

```r
library(SHAPforxgboost)
library(xgboost)
library(data.table)
library(ggplot2)

make_shap_xgb_dependence = function(X,y, interactions=TRUE){

  mod_xgb <- xgboost::xgboost(data=X, label=y, nrounds = 20)
  shap_values <- shap.values(xgb_model = mod_xgb, X_train = X)
  shap_long <- shap.prep(shap_contrib = shap_values$shap_score, X_train = X)
  if (!interactions){
    shap.plot.dependence(
      data_long = shap_long,
      x="age", y = "treatment", color='treatment')
  } else {
    shap.plot.dependence(
      data_long = shap_long,
      data_int = predict(mod_xgb, X, predinteraction = TRUE),
      x="age", y = "treatment", color='treatment')
  }
}
```

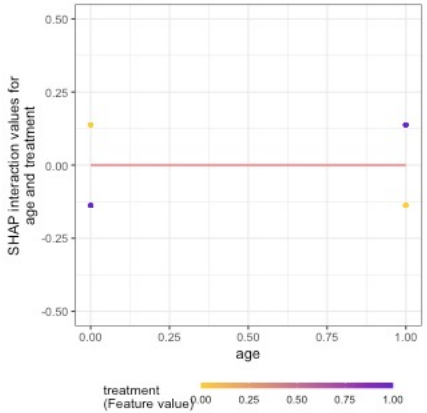
How can we detect an interaction in ML models?

The dependence plot reveals no average effect for age, though larger vertical variation for younger age patients suggests an interaction effect with age.

The interaction plot focuses on the interplay between age and treatment and forces symmetry, which splits that interaction effect in half for both young and older patients, therefore also flipping the direction of the effect as age changes.



Dependence plot

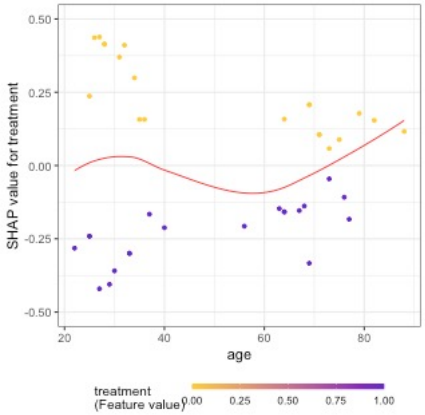


Interaction plot

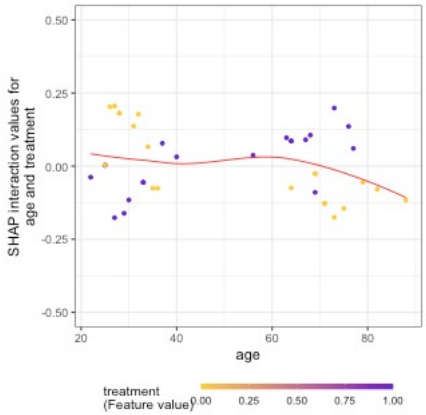
How can we detect an interaction in ML models?

Replacing age with its underlying continuous values yields a similar picture. The dependence plot shows higher vertical dispersion for younger age, suggesting the presence of an interaction effect.

This is confirmed in the interaction plot via the flipping of the colors as we go from younger to older ages.



Dependence plot



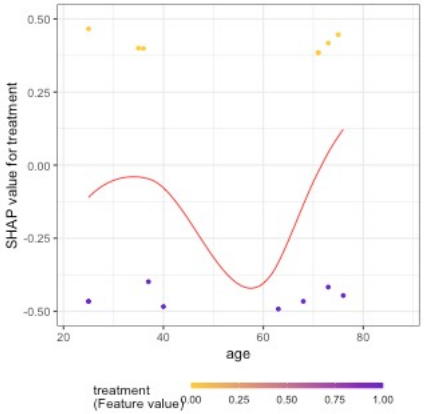
Interaction plot

How can we detect an interaction in ML models?

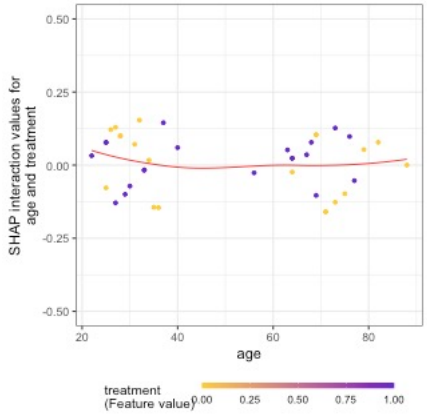
To stress-test our interpretation, let us now fit the same model to the additive dataset, where no interaction effect exists.

Here the degree of variation remains static across different values of age in the dependence plot, whereas in the interaction plot, there is no discernible “flip” of colours across different ages.

This is not completely straightforward.



Dependence plot



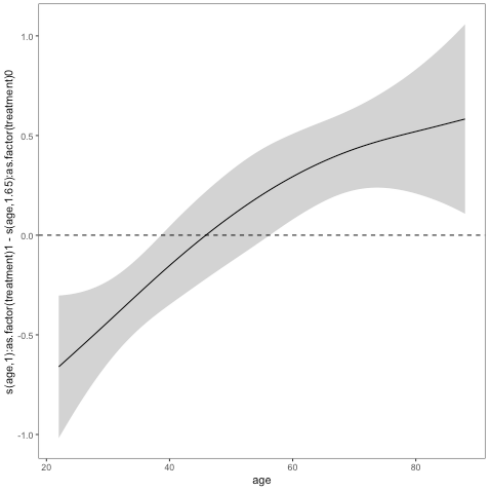
Interaction plot

Interaction terms in GAMs

In GAMs, the shape of the interaction effect does not need to be a linear function of the product of the two variables. It can be an arbitrary 2D smooth. In particular, there are two main approaches:

Binary by continuous	`gam(y ~ s(x, by=z))`
Continuous by continuous	`gam(y ~ s(x, z))`

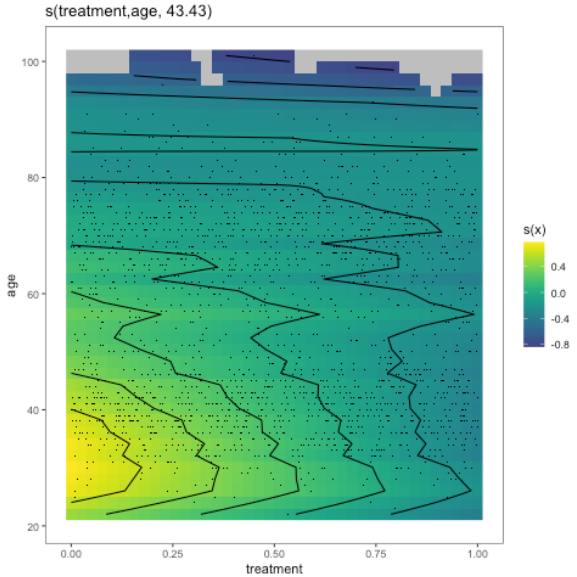
```
library(mgcv)
install.packages('mgcViz')
m_gam = gam(y~ s(age, by=as.factor(treatment)), data=Dint2)
v_gam = getViz(m_gam)
plotDiff(s1 = sm(v_gam, 2), s2 = sm(v_gam, 1)) + l_ciPoly() +
  l_fitLine() + geom_hline(yintercept = 0, linetype = 2)
```



Interaction terms in GAMs

Changing the treatment variable from binary to a uniform in [0,1] (perhaps thought of as dosage) allows us to fit a tensor interaction

```
Family: gaussian
Link function: identity

Formula:
y ~ s(treatment, age, k = 60) + treatment + age

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.047847   0.048426   0.988    0.323
treatment   -0.020097   0.022336  -0.900    0.368
age          0.020848   0.001086  19.202   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                  edf Ref.df     F p-value
s(treatment,age) 43.43   52.2 30.31  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
v_gam = getViz(m_gam)
plot(sm(v_gam, 1)) + l_fitRaster() + l_fitContour() + l_points()
```

Conclusion

- Interaction terms are effects that violate the property of additivity across 2 or more variables.
- The presence of interaction effects is a concern of a different nature than non-linearity.
- Decision trees very naturally represent interactions, but have no way of parsimoniously representing additivity, which makes it harder to extract specific insights about interactions.
- SHAP interactions usefully complement SHAP dependence plots when it comes to interactions. However, they rely too much on “eye-balling” plots, which can be subjective and prone to bias.
- GAMs can only test for pre-specified interactions, rather than discover interactions among all possible candidates, but they offer very precise quantitative insights on interactions.
- The distinction between additive and interacting variables can be critical in some domains. One example is the distinction between predictive and prognostic biomarkers in biostatistics.