36-460/660 Final Project Report

Alex Cheng, Melody Wang, Liz Chu, Kevin Ren

April 23rd, 2024

Contents

Run, Run, then Run Again **1**

Introduction **1**

Data **2**

Methods **2**

Logistic Regression 2

General Additive Model 4

Multimonial Regression 6

Multilevel Regression 7

Results **8**

Discussion **8**

Run, Run, then Run Again

Introduction

The problem of playcalling is among the greatest challenges in sports coaching, particularly in the sport of American football. Particularly poor calls – e.g. the Seahawk's decision to pass rather than run with 1 yard to go in the closing seconds of Super Bowl XLIX – represent snap decisions that can cause franchises and players alike to lose critical games and ultimately millions of dollars. As such, in order to avoid such outcomes coaches must be very smart about what kinds of plays to run in different scenarios, depending on the previous plays that they have called along with the context of the game. In this project we hope to answer the age-old question of what plays a football coach should call, given the types of the previous three plays in the drive, along with contextual information about the field position such as yards-to-go until a touchdown, the strength of the offensive team in terms of their passing and rushing abilities, and the difference in score in the game. We ultimately find that our fitted models on the given covariates perform strongly on both in-distribution and out-of-distribution data on their respective predictive tasks, compared to a naive baseline model.

Data

Describes the data you're using in detail, where you accessed it, along with relevant exploratory data analysis (EDA). You should also include descriptions of any major pre-processing steps.

TODO: we should probably include the code that got us runrunrun.csv somewhere in our code submission

Methods

Describes the modeling techniques you chose, their assumptions, justifications for why they are appropriate for the problem, and your plan for comparison/evaluation approaches. Additionally, you will need to describe how you will quantify uncertainty for your estimates of interest, with sufficient descriptions of the approach and justification for why it's appropriate for your data and problem of interest.

To link our data to our curiosity of playcalling strategy, we focused on 4 methods: Logistic Regression, Generalized Additive Modeling, Multimonial Modeling, and Multilevel Modeling.

Logistic Regression

As we assume a linear relationship among our covariates, a logistic regression was a relatively straightforward choice given that the nature of our question involves classifying the next best play, which is a classification problem.

```
##
## Call:
## glm(formula = outcome ~ as.factor(prev_play_1) + as.factor(prev_play_2) +
##     as.factor(prev_play_3) + ydstogo + score_differential + pass_rank +
##     rush_rank, family = "binomial", data = success_df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.661  -1.014  -0.782   1.212   2.790
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 0.914278   0.077839  11.746  < 2e-16 ***
## as.factor(prev_play_1)pass -0.293999   0.062338  -4.716 2.40e-06 ***
## as.factor(prev_play_1)run  -0.371825   0.062708  -5.929 3.04e-09 ***
## as.factor(prev_play_2)pass -0.545162   0.065173  -8.365  < 2e-16 ***
## as.factor(prev_play_2)run  -0.401489   0.064313  -6.243 4.30e-10 ***
## as.factor(prev_play_3)pass  0.315917   0.056155   5.626 1.85e-08 ***
## as.factor(prev_play_3)run   0.593390   0.057598  10.302  < 2e-16 ***
## ydstogo                    -0.119269   0.005033 -23.698  < 2e-16 ***
## score_differential          0.019753   0.001770  11.163  < 2e-16 ***
## pass_rank                   0.014324   0.001930   7.421 1.16e-13 ***
## rush_rank                  -0.014433   0.001882  -7.669 1.74e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 20195  on 14959  degrees of freedom
## Residual deviance: 19222  on 14949  degrees of freedom
## AIC: 19244
##
## Number of Fisher Scoring iterations: 4
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.9143	0.07784	11.75	7.423e-32
as.factor(prev_play_1)pass	-0.294	0.06234	-4.716	2.403e-06
as.factor(prev_play_1)run	-0.3718	0.06271	-5.929	3.039e-09
as.factor(prev_play_2)pass	-0.5452	0.06517	-8.365	6.017e-17
as.factor(prev_play_2)run	-0.4015	0.06431	-6.243	4.299e-10
as.factor(prev_play_3)pass	0.3159	0.05615	5.626	1.846e-08
as.factor(prev_play_3)run	0.5934	0.0576	10.3	6.885e-25
ydstogo	-0.1193	0.005033	-23.7	3.763e-124
score_differential	0.01975	0.00177	11.16	6.217e-29
pass_rank	0.01432	0.00193	7.421	1.163e-13
rush_rank	-0.01443	0.001882	-7.669	1.737e-14

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	20195 on 14959 degrees of freedom
Residual deviance:	19222 on 14949 degrees of freedom

While all of our coefficients estimates are significant, it seems like the play types of the 3 previous plays brought the most influence on outcome in terms of magnitude. In specific, the 3rd previous play yielded the most change in log-odds of a successful game outcome - the change in the game outcome when the 3rd previous play was a pass compared to when it was the base condition (first_play) was 0.3159 while it was 0.5934 for a run. Since the log odds for the 1st and 2nd previous play was negative, we convert them to an exponential scale and discover that for the 1st previous play, a pass had a higher odds ratio of 0.7452765 than that of the a run. On the second play, however, a pass produced a lower odds ratio of 0.5797258 than the 0.6693153 yielded by a run. Interestingly, it seems like the odds ratio for both types of plays does not necessarily decrease or increase as k changes. More so, yards-to-go, score differential, play_rank and pass_rank seem to have more of an influence on outcome than the previous 2nd and 3rd plays. This may signify that we need to work with more intricate relationships and compare these results with the results of more complex models.

Table 3: Table continues below

(Intercept)	as.factor(prev_play_1)pass	as.factor(prev_play_1)run
2.495	0.7453	0.6895

Table 4: Table continues below

as.factor(prev_play_2)pass	as.factor(prev_play_2)run
0.5797	0.6693

as.factor(prev_play_3)pass	as.factor(prev_play_3)run
1.372	1.81

```
## [1] "run"        "pass"       "first play"
```

```
## [1] 1.371493
```

```
## [1] 1.810132
```

```
## [1] 0.7452765
```

```
## [1] 0.6894921
```

```
## [1] 0.5797258
```

```
## [1] 0.6693153
```

```
## [1] 0.8875415
```

```
## [1] 1.019946
```

```
## [1] 1.014423
```

```
## [1] 0.9856736
```

General Additive Model

Our logistic regression developed promising results, but did not necessarily account for the possibility that we had a non-linear relationship between our predictors and game outcome. By allowing us to smooth over predictors that may have non-linear relationships, we can take advantage of general additive models to flexibly model these intricate relationships.

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## outcome ~ as.factor(prev_play_1) + as.factor(prev_play_2) + as.factor(prev_play_3) +
##     s(ydstogo) + s(score_differential) + pass_rank + rush_rank
##
## Parametric coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -0.194516   0.062638  -3.105   0.0019 **
## as.factor(prev_play_1)pass -0.086438   0.064064  -1.349   0.1773
## as.factor(prev_play_1)run  -0.067655   0.065963  -1.026   0.3051
## as.factor(prev_play_2)pass -0.649021   0.066413  -9.772  < 2e-16 ***
## as.factor(prev_play_2)run  -0.501338   0.065565  -7.646 2.07e-14 ***
## as.factor(prev_play_3)pass  0.326274   0.057348   5.689 1.28e-08 ***
## as.factor(prev_play_3)run   0.596848   0.058698  10.168  < 2e-16 ***
```

```
## pass_rank                  0.014828   0.001960   7.566 3.84e-14 ***
## rush_rank                 -0.014406   0.001916  -7.520 5.47e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                         edf Ref.df Chi.sq p-value
## s(ydstogo)            6.473  7.304  863.3  <2e-16 ***
## s(score_differential) 6.017  7.045  159.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0884   Deviance explained =  6.8%
## UBRE = 0.26105  Scale est. = 1         n = 14960
```

Odds Ratio:

Table 6: Table continues below

(Intercept)	as.factor(prev_play_1)pass	as.factor(prev_play_1)run
0.8232	0.9172	0.9346

Table 7: Table continues below

as.factor(prev_play_2)pass	as.factor(prev_play_2)run
0.5226	0.6057

as.factor(prev_play_3)pass	as.factor(prev_play_3)run
1.386	1.816

Table 9: Table continues below

Intercept	prev_1_pass	prev_1_run	prev_2_pass	prev_2_run	prev_3_run
1.215	0.9172	0.9346	0.5226	0.6057	1.386

prev_3_pass	pass_rank	push_rank
1.816	1.015	0.9857

Surprisingly, after smoothing for score_differential and yds_to_go, the previous 1st play type lost its significance while the ratio for both play types increased substantially. Additionally, the run had a higher odds ratio than that of a pass for the first previous play - the exact opposite as our previous model. However, we developed similar results as logistic regression for the previous second and third play types. We hypothesize that the significant change in the first previous play stems from ydstogo and score differential having a sizable interaction effect on the first previous play, which makes sense as ydstogo and score differential measure current conditions of the game, and the most recent previous play contributes more change to current situations than further plays in time.

Multimonial Regression

We proceed in experimenting with model performance with multimonial modeling, which allows the model to perform with more granularity while still maintaining the form of logistic regression. In specific, we have the presence of short/deep and left/middle/right passes and left/middle/right runs, so a multimonial model would be suitable for interpreting the effect on predictors of multiple categories on the outcome.We first filter out the NAs in the pass types, as there are more null values in our more granular predictors.

Table 11: Table continues below

	(Intercept)	as.factor(prev_play_1)pass
pass deep middle	0.4199	0.9244
pass deep right	1.392	0.8227
pass short left	9.649	1.082
pass short middle	4.126	1.325
pass short right	7.944	1.013
run left	20.81	0.8547
run middle	26.84	0.6685
run right	17.08	0.8263

Table 12: Table continues below

	as.factor(prev_play_1)run	as.factor(prev_play_2)pass
pass deep middle	0.9406	0.8115
pass deep right	0.7093	1.125
pass short left	1.026	0.5688
pass short middle	1.106	0.6352
pass short right	0.994	0.5623
run left	0.7039	0.3759
run middle	0.5594	0.4055
run right	0.7711	0.3279

Table 13: Table continues below

	as.factor(prev_play_2)run	as.factor(prev_play_3)pass
pass deep middle	1.117	1.129
pass deep right	1.089	1.347
pass short left	0.5533	1.724
pass short middle	0.5895	1.65
pass short right	0.5718	1.921
run left	0.4233	2.274
run middle	0.4291	2.334
run right	0.4051	2.284

Table 14: Table continues below

	as.factor(prev_play_3)run	ydstogo
pass deep middle	1.1	1.043
pass deep right	1.188	0.9821

	as.factor(prev_play_3)run	ydstogo
pass short left	1.64	0.9326
pass short middle	1.502	0.9502
pass short right	1.822	0.9367
run left	2.996	0.8578
run middle	2.548	0.7782
run right	2.916	0.8661

	score_differential	pass_rank	rush_rank
pass deep middle	0.99	0.9817	1.006
pass deep right	1.005	0.9916	1.004
pass short left	0.9994	1.004	0.9977
pass short middle	0.9955	1.003	1.001
pass short right	1.002	1.005	1.009
run left	1.022	1.011	0.9875
run middle	1.015	1.017	0.9978
run right	1.021	1.023	0.9835

Pass short middle was the best-outcome strategy in term of odds ratio for all the previous plays with the exception of the 3rd previous play. We do not see a lot of difference in Run directions for the previous plays or for the other predictors, but do note that a middle run has a lower log odds than other directions for score_differential and ydstogo. This is reasonable - while a middle run may bring you closer to the other side of the field, the linebackers and nose tackles on defense often clog up the middle lane, which minimizes score potential. Middle runs also have lower odds ratio more often than not in previous plays. [insert myabe more anaalysis on pass types]

Multilevel Regression

Since our effects may not be constant, we let go of the fixed effects assumption present in the previous models to create a multilevel model. Specifically, we take advantage of the model's suitability for nested structures in accounting for dependencies among the same cluster. We also included random intercepts for the offensive and defensive teams and their interaction with each other.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.028	1.097	2061	1
as.factor(prev_play_1)pass	0.7489	1.065	0.009967	1
as.factor(prev_play_1)run	0.6967	1.065	0.003244	1
as.factor(prev_play_2)pass	0.5785	1.068	0.0002338	1
as.factor(prev_play_2)run	0.6733	1.067	0.002212	1
as.factor(prev_play_3)pass	1.342	1.058	183.6	1
as.factor(prev_play_3)run	1.78	1.06	20689	1
ydstogo	0.8875	1.005	5.974e-11	1
pass_rank	1.01	1.003	18.95	1.003

The multilevel model generally follows that of logistic regression in that the 3rd previous play had the highest effect and the 2nd previous play had the lowest.

Results

Describes your results. This can include tables and plots showing your results, as well as text describing how your models worked and the appropriate interpretations of the relevant output. I do not want to you to write out the textbook interpretations of all model coefficients! I only want you to interpret the output that is relevant for your question of interest that is framed in the introduction.

Discussion

Overall it was very interesting to see how similarly our chosen models performed given the breadth of the features we chose. This may suggest that our features were not distinct enough from each other or we did not select enough features. This could be plausible as play-calling is a clearly complex process representing the battle between offensive and defensive coaching minds along with the players on the field, not to mention the field position of the play itself. As such, it is plausible that one limitation of our approach is simply that we did not perform enough vetting of the features we chose. In the future we could potentially add more features such as the ranking of the defensive team on the field versus the pass or the run (currently no defensive statistics are considered, which could be troublesome) as well as including features such as which team is home versus away, or perhaps tuning our models on different number of prior plays that the model is able to see.

Secondly, the question of what our training data should look like was something that we could potentially change in the future as well. By nature our data fails to reveal counterfactual outcomes, e.g. the offensive team induces a treatment in the form of a play call, causing some outcome in the form of yards gained, which doesn't allow for seeing what *would* have happened should a different play have been run. As a result we chose not to include what we determined to be 'failed' plays in our training process, and so we modeled the probability that a given play occurred given that it was successful. This process therefore wastes lots of data: no information from the failed plays is reflected in our trained models. Expanding our analysis to be a causal inference question in which we attempt to do some modeling on what play calls could potentially reverse or alter the outcomes of failed plays could represent an interesting statistical question under which future research could be conducted.