Modeling Marathon Paces

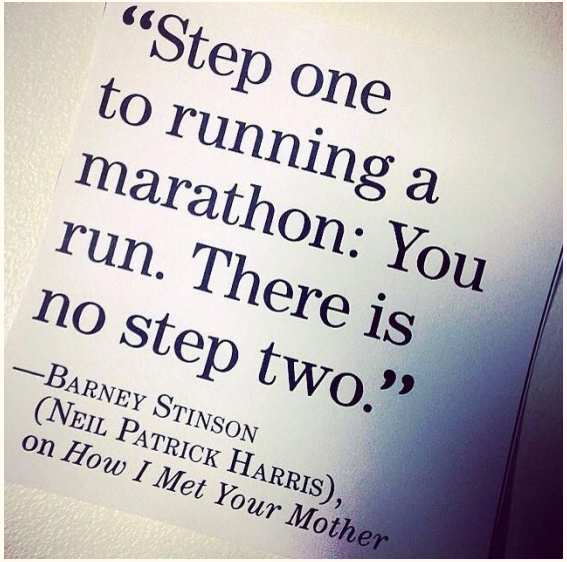
Julia Hinds

CMSE 801-001 Final Project

December 8 , 2021

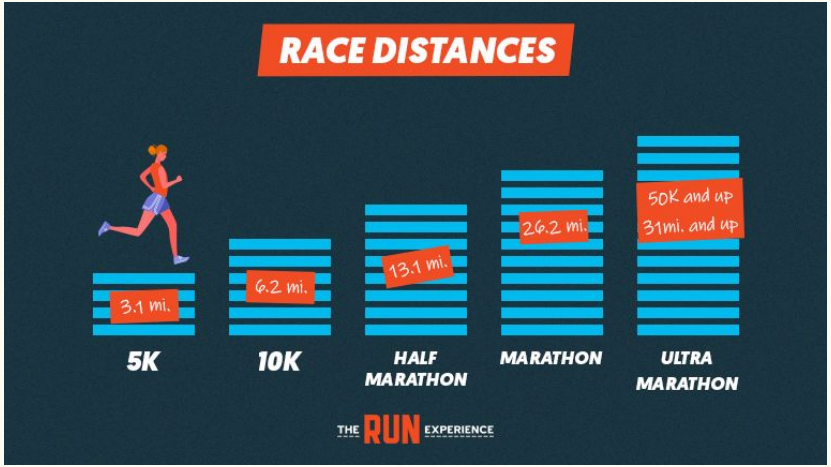
Overview

- Background/Motivation
- Research Question
- Code & Documentation
- Methodology
- Results & Discussion
- Validation
- Limitations & Next Steps
- Conclusion



Background

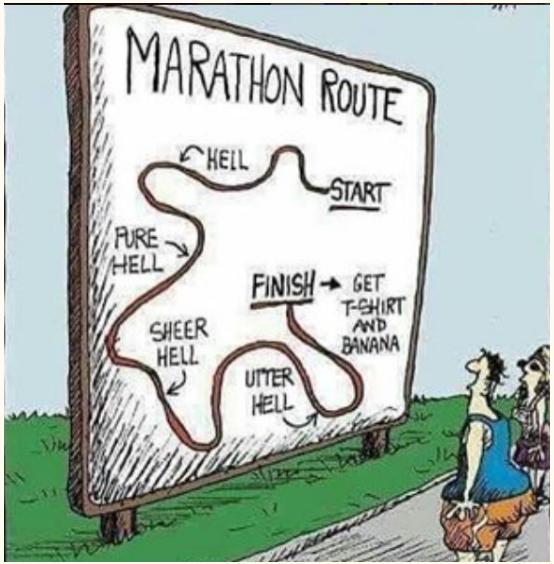
- Length of marathon: 26.2 miles (42.2 km)
- About 1.1 million runners complete a marathon per year[1]
  - Based on 2018 statistics
- Personal qualifications
  - 12 marathons
  - 9 states
  - 1 international
  - 2 ultramarathons
  - Next race: 2022 Kentucky Derby Marathon



1. https://www.livestrong.com/article/13763749-marathon-statistics/
2. Image: https://therunexperience.com/how-far-is-a-marathon-and-should-you-try-to-run-one-jg/

Research Question

Is it possible to accurately predict finish paces during a marathon?



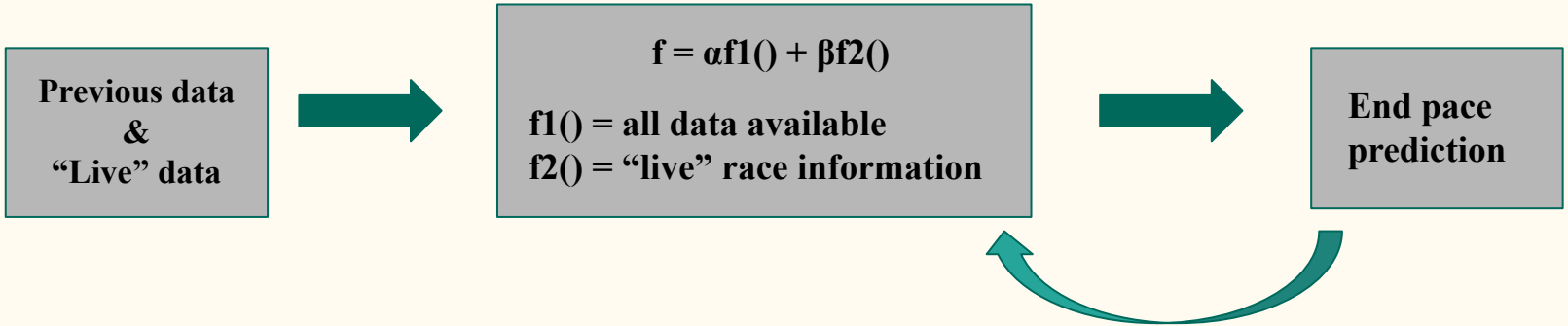
Code & Documentation

- Python (v3.7.1) and Jupyter Notebook were used
- Data: 2015 San Francisco Marathon[1], personal marathon splits
- Python packages used for analysis:
  - Pandas
  - Numpy
  - Sklearn
- Python packages used for visualization
  - Matplotlib

1. https://www.runraceresults.com/Secure/RaceResults.cfm?ID=RCLF2015

Methodology

1. Finding accurate model to model race paces
   a. Sklearn GaussianProcessRegressor
2. “Black box” function to predict end pace

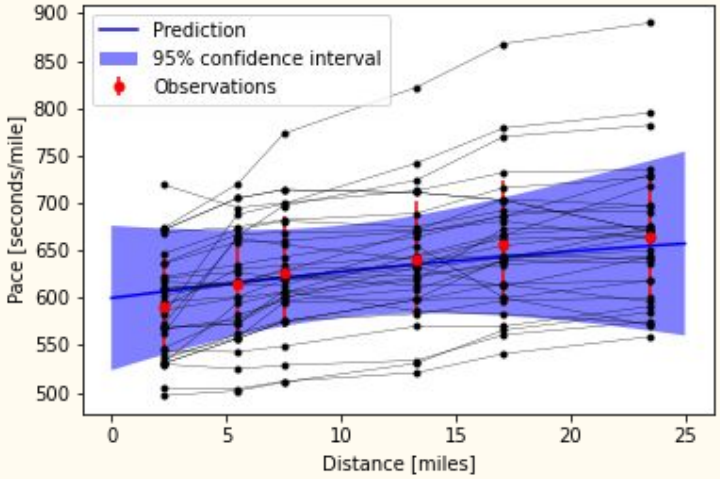


3. Apply to two types of data available
   a. Only polynomial degrees 1 and 3 shown in presentation
   b. Including desired end race pace, without inclusion of desired pace

2015 San Francisco Marathon splits

	Mile	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6	Race 7	Race 8	Race 9	...	Race 25	Race 26	Race 27	Race 28	Race 29	Race 30	Race 31	Race 32	Race 33	Race 34
0	2.3	547	587	573	568	621	596	636	591	672	...	617	505	568	530	610	497	544	535	531	531
1	5.5	582	600	663	573	627	601	659	630	705	...	614	504	573	525	615	502	543	562	557	557
2	7.5	597	606	656	600	620	619	660	634	714	...	613	512	607	529	622	511	549	576	575	575
3	13.3	609	621	640	632	624	631	670	643	711	...	655	521	586	534	613	531	570	583	598	598
4	17.1	639	639	645	612	635	642	681	662	703	...	600	541	583	561	614	566	570	597	600	615

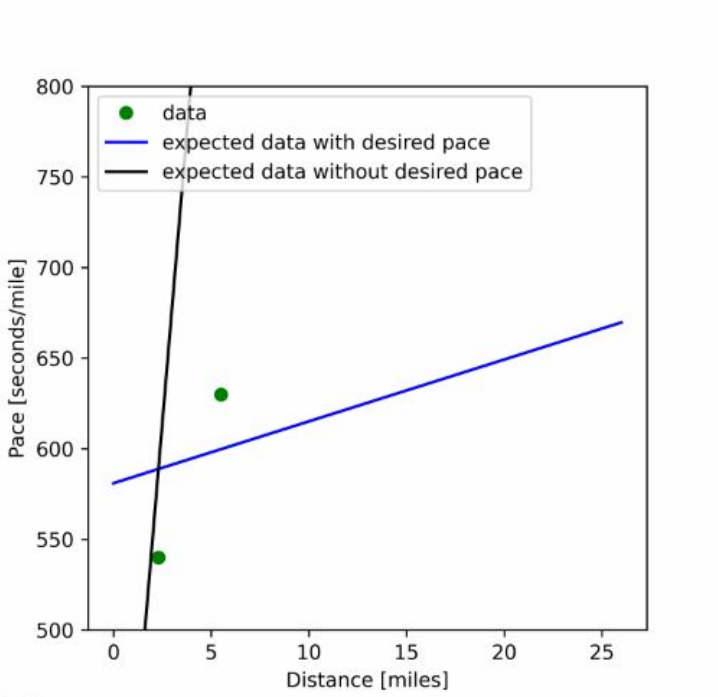
- Pandas.head() used to obtain top table
- Splits are record at 2.3, 5.5, 7.5, 13.3, and 17.1 miles
- Sklearn's GaussianProcessRegressor
- kernel = C(1.0, (1e-3, 1e8)) * RBF(1, (1e-2, 1e2))



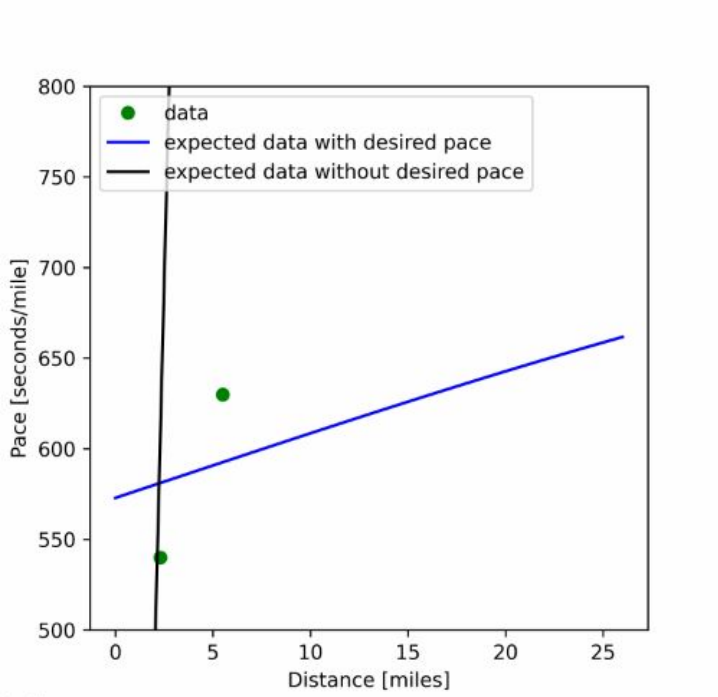
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1



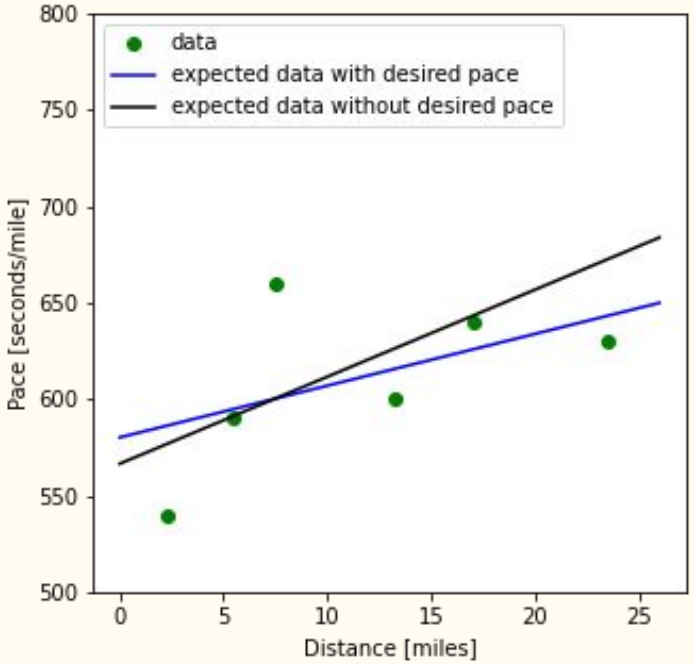
D = 3



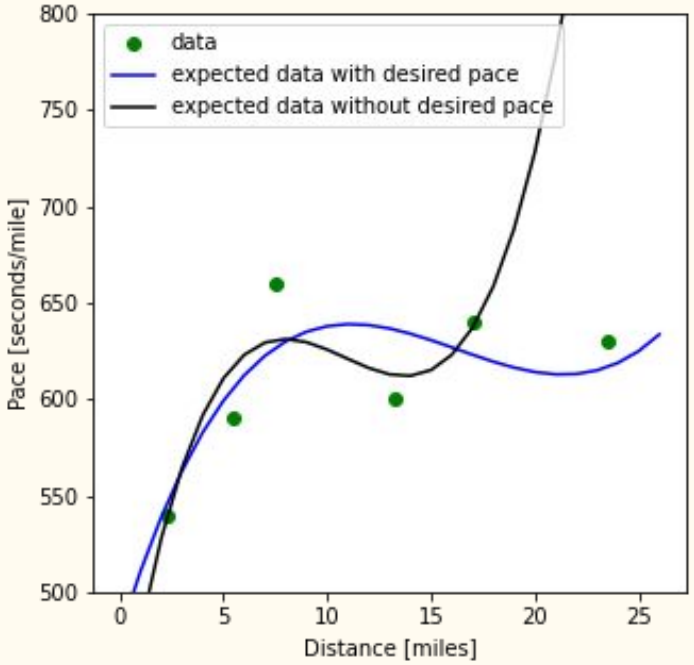
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 640.1 seconds/mi
Projected end-pace based on just data: 677.4seconds/mi



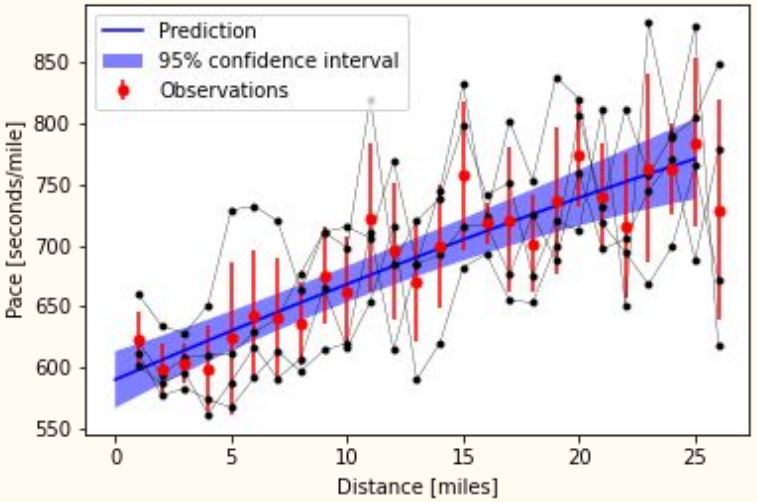
D = 3
Projected end-pace based on goal: 633.8 seconds/mi
Projected end-pace based on just data: 1236.9 seconds/mi



Personal marathon splits

	Mile	Race 1	Race 2	Race 3	Race 4
0	1	601	660	621	612
1	2	587	634	594	577
2	3	596	628	608	583
3	4	562	651	610	574
4	5	588	729	611	567

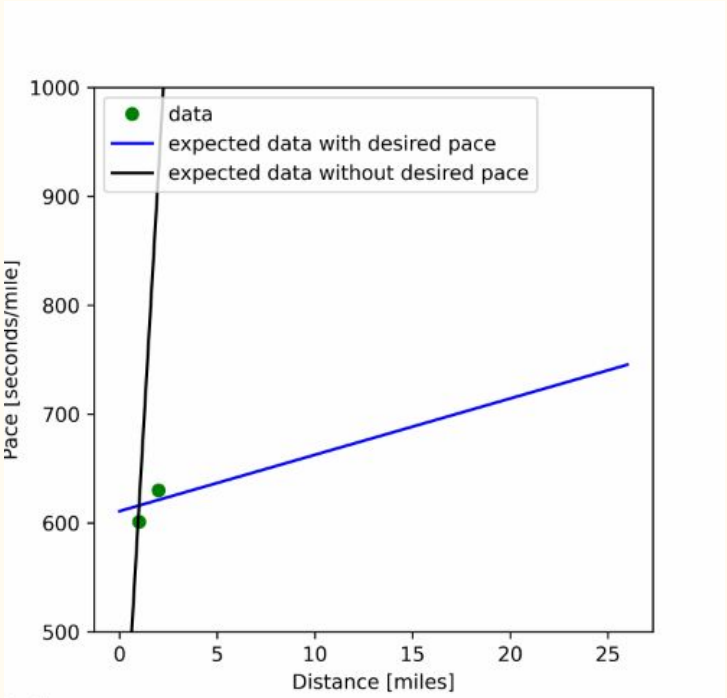
- Pandas.head() used to obtain top table
- Splits are record at every mile from 1 to 26
- Sklearn's GaussianProcessRegressor
- kernel = C(1.0, (1e-3, 1e8)) * RBF(1, (1e-2, 1e2))



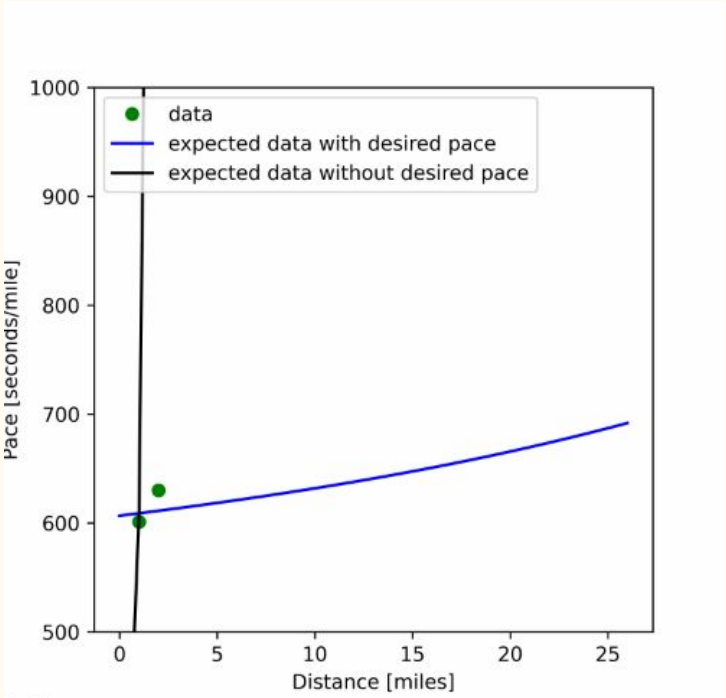
Personal splits animation

Input final desired pace: 630 seconds/mi

D = 1



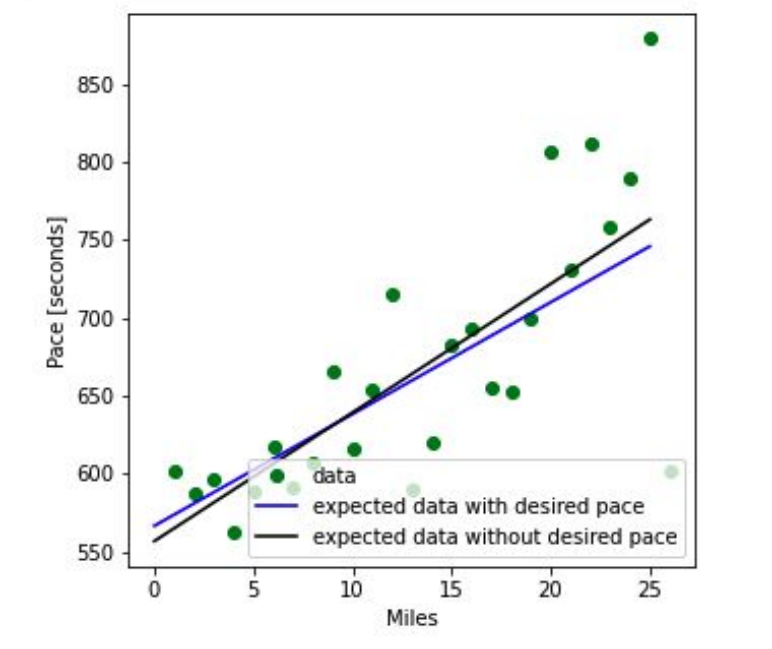
D = 3



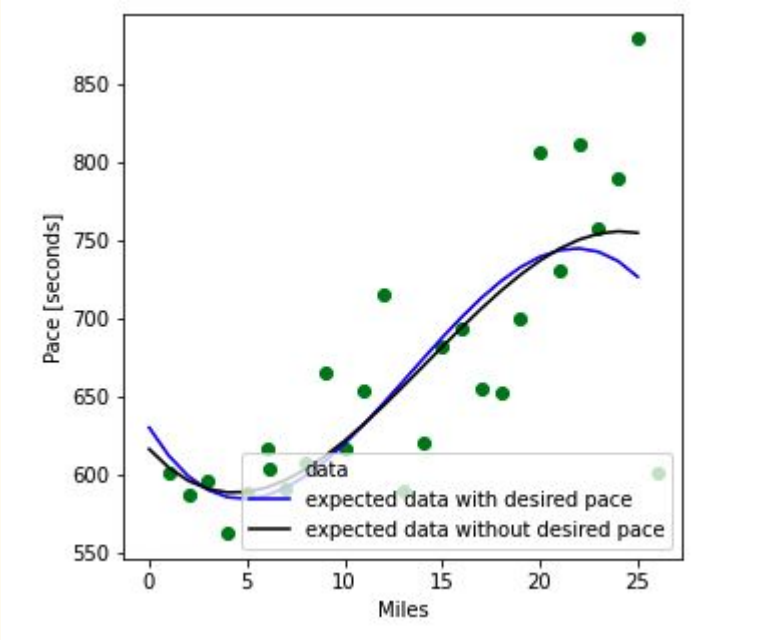
Personal splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 752.9 seconds/mi
Projected end-pace based on just data: 711.2 seconds/mi

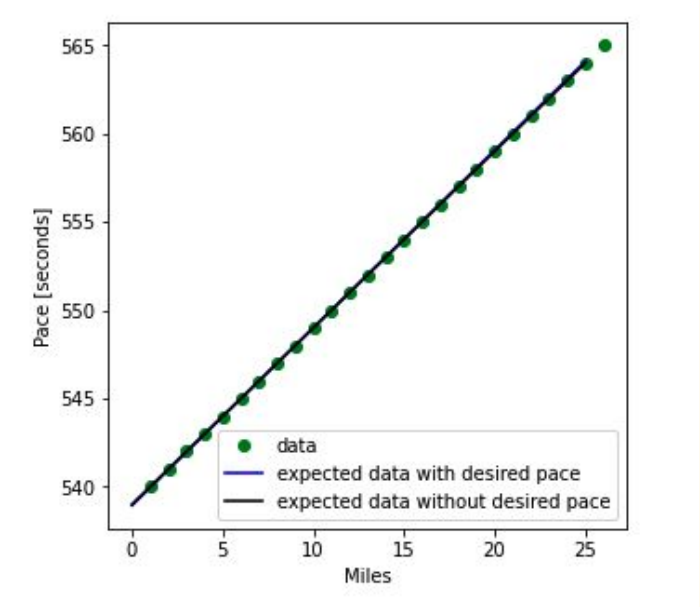


D = 3
Projected end-pace based on goal: 712.0 seconds/mi
Projected end-pace based on just data: 750.9 seconds/mi



Validation

- Created fake “perfect pace” data that is linear



Projected end-pace based on goal: 565 seconds/mi
Projected end-pace based on just data: 565 seconds/mi

Limitations & Next Steps

- Limitations:
  - Need more data!
    - More runners with more splits
      - Races usually don't provide a lot of splits
      - My splits are limited to what I've tracked
    - More physiological variation
    - More races with different terrain
  - Previous training methods, injuries, weather conditions, etc. were not taken into account
- Next Steps
  - Use regression analysis models
  - Determine final race time rather than just final pace
    - Nike Breaking2 project

Conclusion

- It’s difficult to find an accurate model with little data
- With this code, you can *roughly* determine final race pace
- Regression should be used for future studies

Questions?

Backup

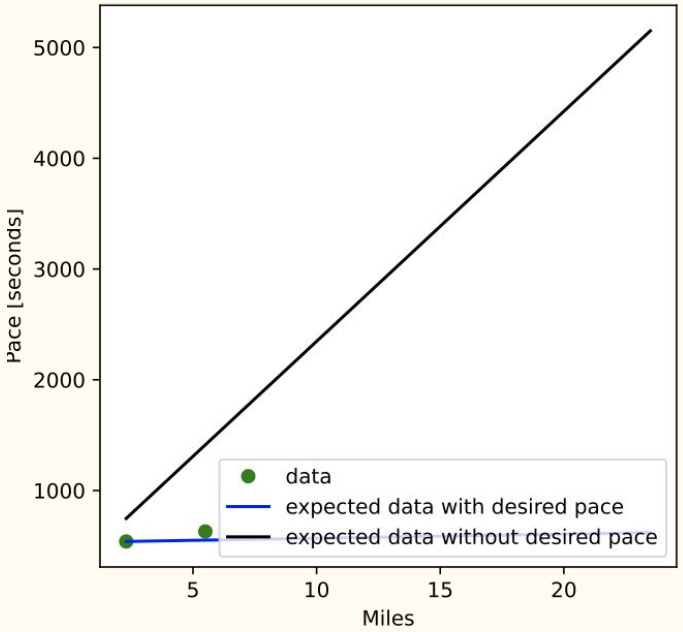
Sklearn GaussianProcessRegressor

- Gaussian process regression (GPR)
  - Probabilistic supervised machine learning framework
  - Makes predictions by incorporating prior knowledge (known as “kernels”) and provide an uncertainty over the predictions
- Kernel (covariance function)
  - Specify the statistical relationship between two points (x_1,x_2) in the input space
  - ConstantKernal
    - $k(x_1+x_2) = constant\_value \ \forall \ x_1,x_2$
  - RBF (Radial-basis function or Squared Exponential Kernel)
    - $k(x_1+x_2) = \sigma^2 exp(- d(x_1,x_2)^2 / 2l^2)$
    - Where l is length scale of kernel, σ^2 is the output variance (essentially a scale factor) and d(.,.) is the Euclidean distance

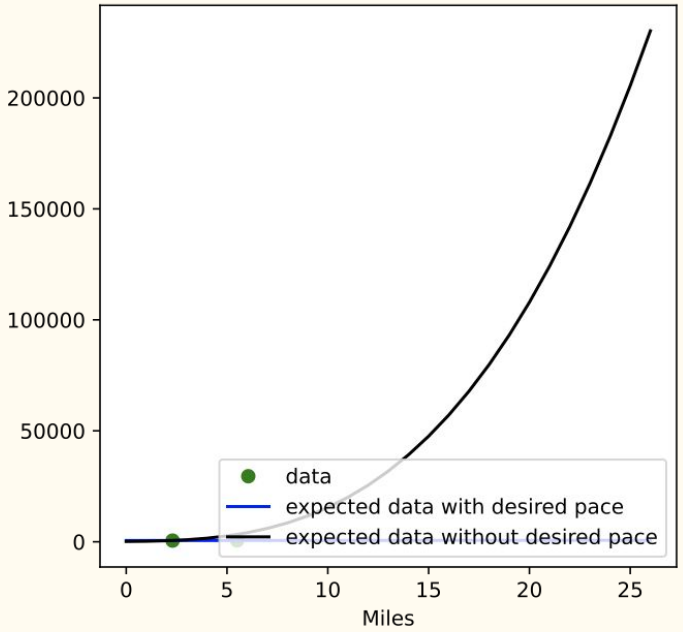
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 619.8 seconds/mi
Projected end-pace based on just data: 5150.8 seconds/mi



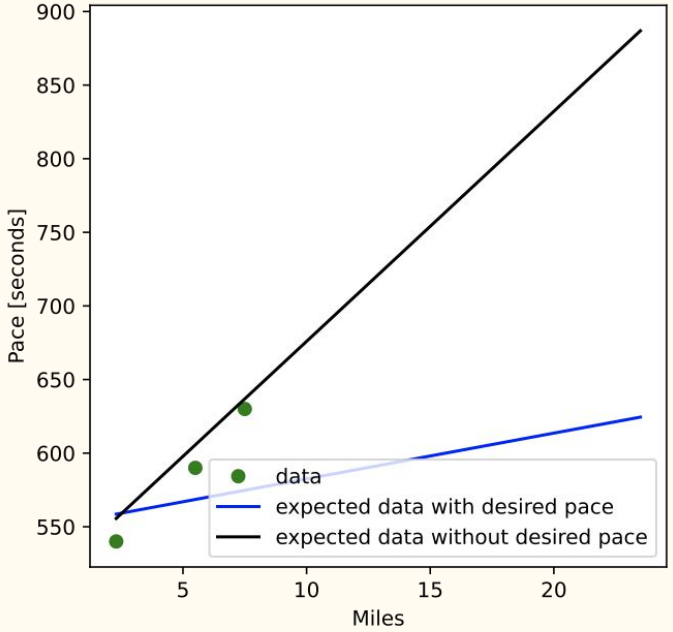
D = 3
Projected end-pace based on goal: 661.7 seconds/mi
Projected end-pace based on just data: 230223 seconds/mi



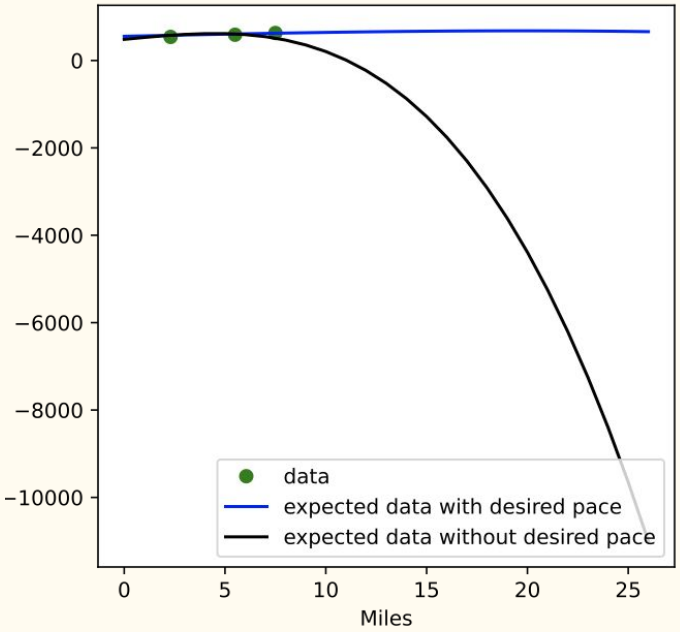
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 624.5 seconds/mi
Projected end-pace based on just data: 886.9 seconds/mi



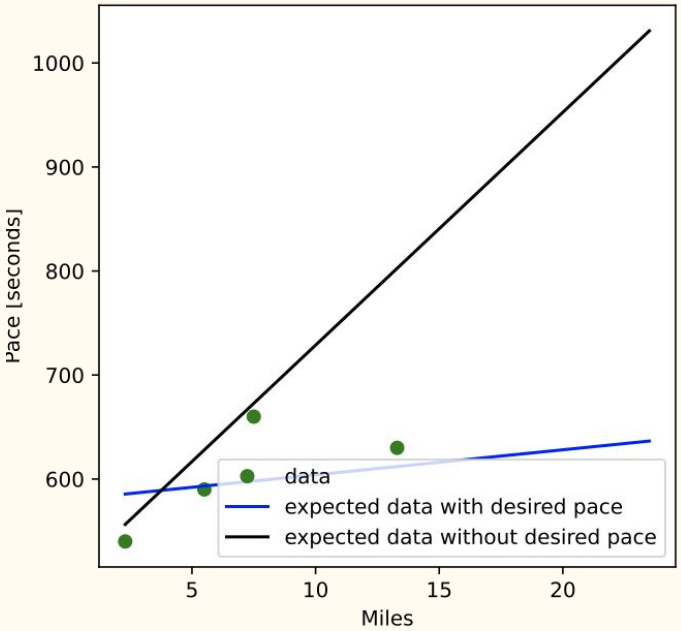
D = 3
Projected end-pace based on goal: 660.5 seconds/mi
Projected end-pace based on just data: -11008 seconds/mi



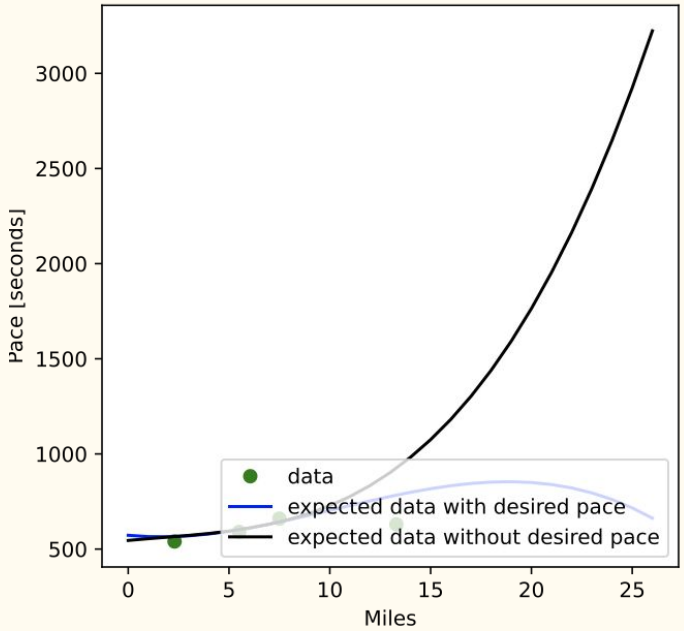
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 636.5 seconds/mi
Projected end-pace based on just data: 1030.9 seconds/mi



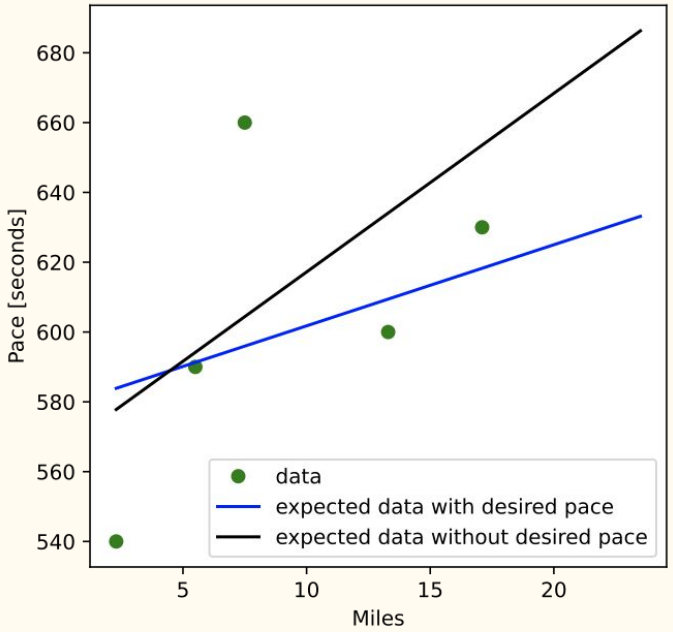
D = 3
Projected end-pace based on goal: 662.2 seconds/mi
Projected end-pace based on just data: 3223 seconds/mi



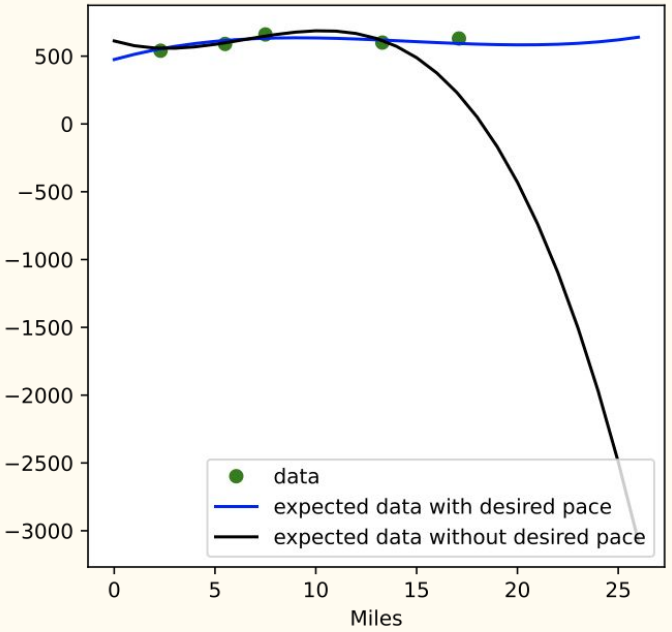
SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 633.1 seconds/mi
Projected end-pace based on just data: 686.3 seconds/mi



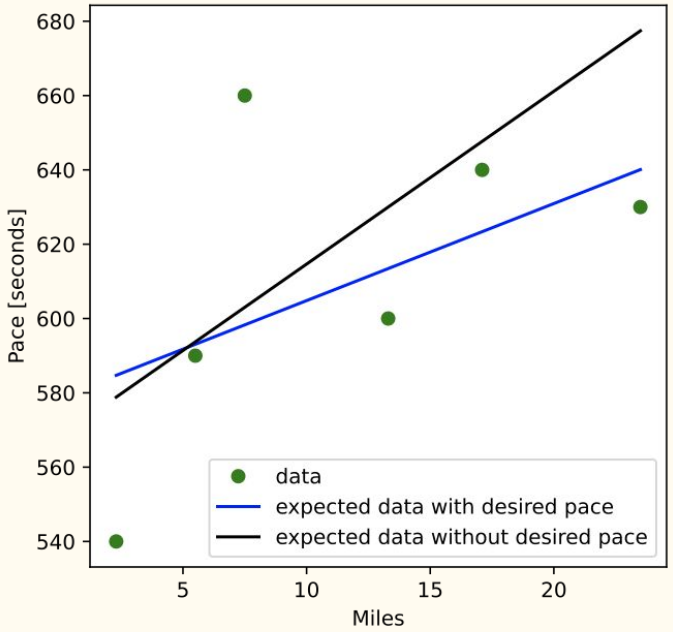
D = 3
Projected end-pace based on goal: 639.6 seconds/mi
Projected end-pace based on just data: -3078 seconds/mi



SF Marathon splits animation

Input final desired pace: 630 seconds/mi

D = 1
Projected end-pace based on goal: 640.1 seconds/mi
Projected end-pace based on just data: 677.4seconds/mi



D = 3
Projected end-pace based on goal: 633.8 seconds/mi
Projected end-pace based on just data: 1236.9 seconds/mi