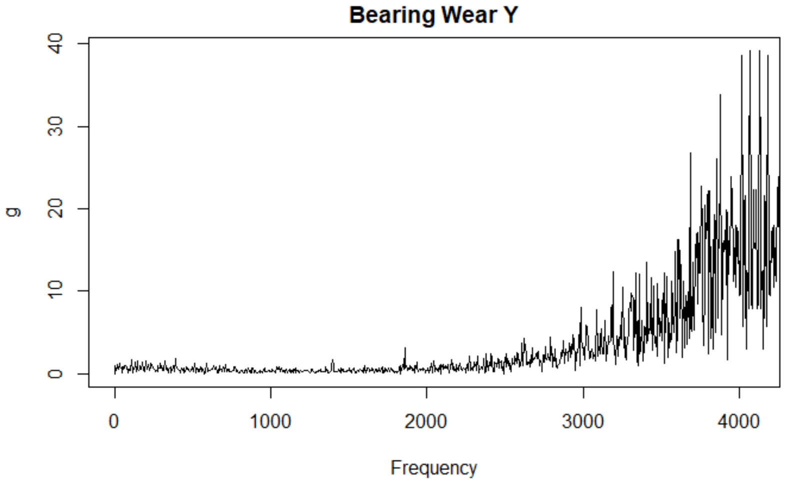
# Data Analytics Engineer Case Study

Aaron Kruchten

April 17, 2021

# Bearing Wear FFT Acceleration Y Axis



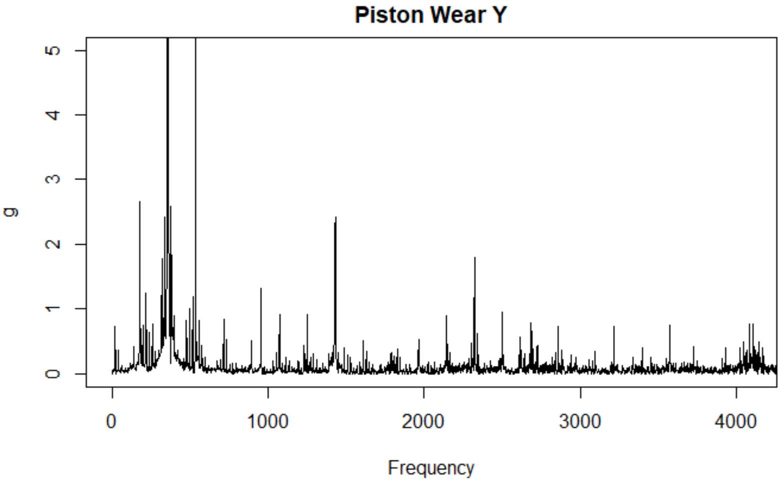
# Bearing Wear FFT Acceleration Y Axis

- Use "Weighted Average of Frequencies of Peaks" as feature for this failure state.
- To do this, first scale each data down to the interval $[0, 1]$ by dividing each data point by the sum of all data points.
- $|x_i|$ is the euclidean norm of a complex number output by the discrete fourier transform. $i \in [N] \cup \{0\}$
- $\mu$ and $\sigma^2$ are the mean and variance of the scaled data respectively.
- $\mathbb{I}_{Peak,i}$ is an indicator which is 1 if $|x_i| > \theta_1\mu + \theta_2\sigma^2$ and 0 otherwise.
- $\sum_{i=0}^{N/2} i * \frac{8192}{N} * \mathbb{I}_{Peak,i} * |x_i|$

# Bearing Wear FFT Acceleration Y Axis

- One issue with generalization is that we could have one peak near where the weighted average is computed for Bearing Wear.
- Could fix this issue by also counting the number of peaks.
- Second, we could have two groups of peaks which give us the same average.
- Could fix this issue by counting the number of groups of peaks. This would likely involve using a rolling variance.

# Piston Wear Acceleration Y Axis



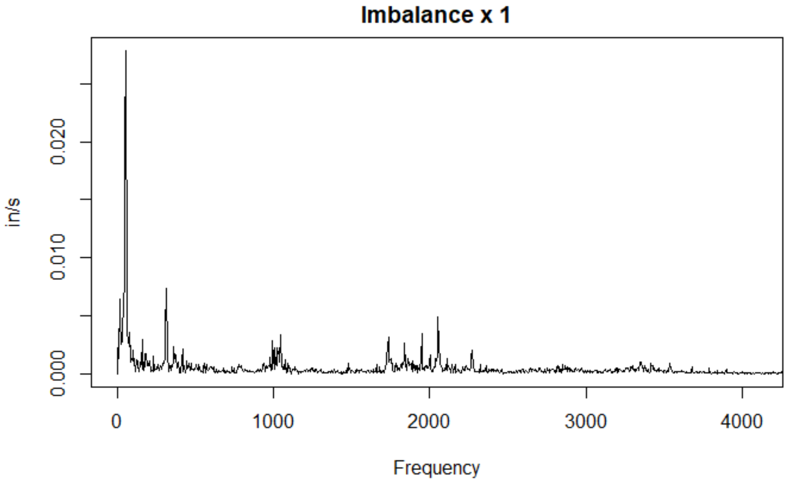
# Piston Wear Acceleration Y Axis

- Use the variance of the distance between peaks as feature for this failure state.
- $\mu_{20}$ and $\sigma_{20}^2$ are the rolling mean and rolling variance respectively.
- Peaks are classified when $|x_i| > \theta_1 \mu_{20} + \theta_2 \sigma_{20}^2$
- Then, compute the difference in frequencies between each of the adjacent peaks. Then compute the variance of these differences in frequencies.
- Variance of these differences should be lower for this failure state than any of the others.

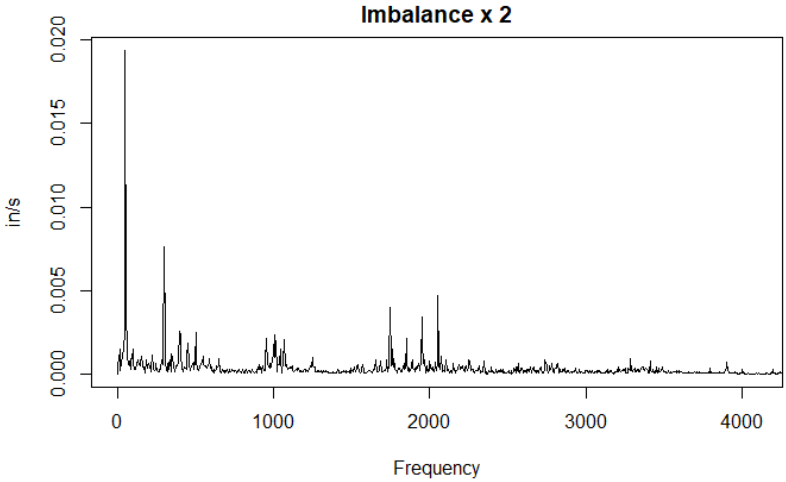
# Piston Wear Acceleration Y Axis

- This should generalize well as long as there is no other failure state which has periodic peaks. If so we could also include the average distance between peaks to try and differentiate this failure condition from other signals with periodic peaks.
- This also may not generalize well if we only have a few peaks. For example, if we only have two peaks then the variance between the peaks would be zero. We could fix this by counting the number of peaks.

# Imbalance Velocity X Axis 1

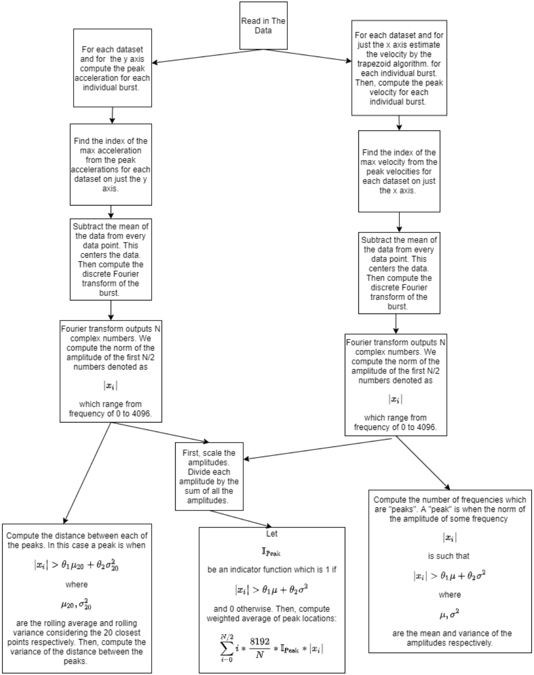


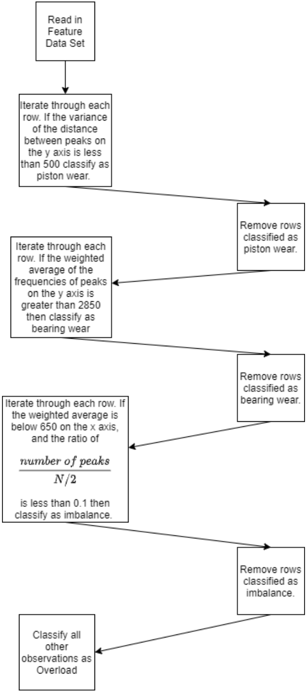
# Imbalance Velocity X Axis 2



# Imbalance Velocity X Axis

- We used the trapezoid algorithm to approximate the integral of acceleration. This gave us velocity.
- To classify this failure state we compute the weighted average as we did with bearing wear, and the number of peaks. We consider a data point a peak in the same way as we have in the previous two cases.


Read in The Data
For each dataset and for the y axis compute the peak acceleration for each individual burst.
For each dataset and for just the x axis estimate the velocity by the trapezoid algorithm. for each individual burst. Then, compute the peak velocity for each individual burst.
Find the index of the max acceleration from the peak accelerations for each dataset on just the y axis.
Find the index of the max velocity from the peak velocities for each dataset on just the x axis.
Subtract the mean of the data from every data point. This centers the data. Then compute the discrete Fourier transform of the burst.
Subtract the mean of the data from every data point. This centers the data. Then compute the discrete Fourier transform of the burst.
Fourier transform outputs N complex numbers. We compute the norm of the amplitude of the first N/2 numbers denoted as $|x_i|$ which range from frequency of 0 to 4096.
Fourier transform outputs N complex numbers. We compute the norm of the amplitude of the first N/2 numbers denoted as $|x_i|$ which range from frequency of 0 to 4096.
First, scale the amplitudes. Divide each amplitude by the sum of all the amplitudes.
Compute the number of frequencies which are "peaks". A "peak" is when the norm of the amplitude of some frequency $|x_i|$ is such that $|x_i| > \theta_1 \mu + \theta_2 \sigma^2$ where $\mu, \sigma^2$ are the mean and variance of the amplitudes respectively.
Compute the distance between each of the peaks. In this case a peak is when $|x_i| > \theta_1 \mu_{20} + \theta_2 \sigma_{20}^2$ where $\mu_{20}, \sigma_{20}^2$ are the rolling average and rolling variance considering the 20 closest points respectively. Then, compute the variance of the distance between the peaks.
Let $\mathbb{I}_{\text{Peak}}$ be an indicator function which is 1 if $|x_i| > \theta_1 \mu + \theta_2 \sigma^2$ and 0 otherwise. Then, compute weighted average of peak locations: $\sum_{i=0}^{N/2} i * \frac{8192}{N} * \mathbb{I}_{\text{Peak}} * |x_i|$


Read in Feature Data Set
Iterate through each row. If the variance of the distance between peaks on the y axis is less than 500 classify as piston wear.
Remove rows classified as piston wear.
Iterate through each row. If the weighted average of the frequencies of peaks on the y axis is greater than 2850 then classify as bearing wear
Remove rows classified as bearing wear.
Iterate through each row. If the weighted average is below 650 on the x axis, and the ratio of
$\frac{number\ of\ peaks}{N/2}$
is less than 0.1 then classify as imbalance.
Remove rows classified as imbalance.
Classify all other observations as Overload

# Generalizing Bearing Wear

- We want to choose parameters $\theta_1$ and $\theta_2$ which are optimal for classifying a signal which is showing signs of bearing wear.
- If we can express the probability of a signal being bearing wear as a differentiable function of our parameters $\theta_1$ and $\theta_2$ then we can use the principle of maximum likelihood estimation to find optimal parameters.
- Some Notation:
- $\tilde{\theta}$ is a column vector of our parameters $\theta_1, ..., \theta_4$
- $s(x, k) = \frac{1}{1+e^{-k(x)}}$ where $k \geq 1$. This acts as a differentiable indicator function.

# Generalizing Bearing Wear

- Assume we have $M$ data sets which are independent observations, showing signs of bearing wear and some not.
- $|x_i^{(j)}|$ is the $i_{th} \in [\frac{N}{2}] \cup \{0\}$ euclidean norm of the fft of our signal for the $j_{th} \in [M]$ signal.
- $y^{(j)}$ is an indicator which is one if the $j_{th}$ signal is bearing wear and 0 otherwise.
- $x^{(j)}$ is a vector of the euclidean norms of an fft of a signal with $\frac{N}{2}$ numbers.
- $L(\tilde{\theta}|x) = \prod_{j=1}^{M} p(x^{(j)})^{y^{(j)}}(1 - p(x^{(j)}))^{(1-y^{(j)})}$
- $p(x^{(j)})$ is the probability of the $j_{th}$ signal showing signs of bearing wear.

# Generalizing Bearing Wear

- $s(|x_i^{(j)}| - \theta_1\mu - \theta_2\sigma^2, k)$
- $z^{(j)} = \sum_{i=0}^{N/2} i * \frac{8192}{N} * s(|x_i^{(j)}| - \theta_1\mu - \theta_2\sigma^2, k) * |x_i^{(j)}|$. This gives the "weighted average feature" but it is differentiable.
- $p(x^{(j)}) = \frac{1}{1+e^{-(\theta_4+\theta_3 z^{(j)})}}$
- A standard technique is compute the log likelihood. Therefore, we have.
- $log(L(\tilde{\theta}, x)) = \sum_{j=1}^{M} y^j log(p(x^{(j)}) + (1 - y^{(j)})(log(1 - p(x^{(j)}))$

# Generalizing Bearing Wear

- To choose optimal parameters for $\theta_1, ..., \theta_4$ we would compute the partial derivatives of $L(\tilde{\theta}|x)$ with respect to $\theta_1, ..., \theta_4$. Then, set them equal to zero and solve. There is no way to to do this analytically.
- Therefore, we must use some iterative method. An equivalent way to maximizing the log likelihood is taking the negative of the function and minimizing it with some iterative method such as stochastic gradient descent.
- One possible approach for to generalize is to fit a statistical model using each of the $\frac{N}{2}$ $|x_i|$ as features. However, this would require a lot of data to work effectively. My approach only learns four parameters and should work better with small data sets.

# Issues with Generalizing Bearing Wear

- Update Rule for SGD: $\tilde{\theta} \leftarrow \tilde{\theta} - \gamma \Delta_{\tilde{\theta}} L(\tilde{\theta}|x)$ where $\gamma$ small usually $0 < \gamma < 1$
- We defined our function $s(x, k)$ to act like an indicator function. The larger we make $k$ the closer this function will be to an indicator, however this will make the gradient very large. This may cause an error called the exploding gradient problem.
- Exploding gradient problem is when the gradient is so large that the model will essentially skip over the minimum.

# Further Generalizing

- We could do essentially do the same thing for imbalance.
- For piston wear, I believe there is no way to express the variance of the difference in frequencies between peaks as a differentiable function. However, I came up with a heuristic that may allow it to still work.