Outline

- Brief Introduction
- Feature Extraction
- Classification Models: Random Forest & kNN
- Conclusion and Final Thoughts

Introduction

- We looked at tiny little things called Plankton
- These tiny little things like to float around in the sea
- They also like to be eaten

Questions of Interest

- Histogram v. Moments: Dawn of Justice
- Is there an improvement in prediction combining these methods?
- Does a change in the environment effect our predictions?

Kratuchok's Moments

- Calculating *Kratuchok's* moments,

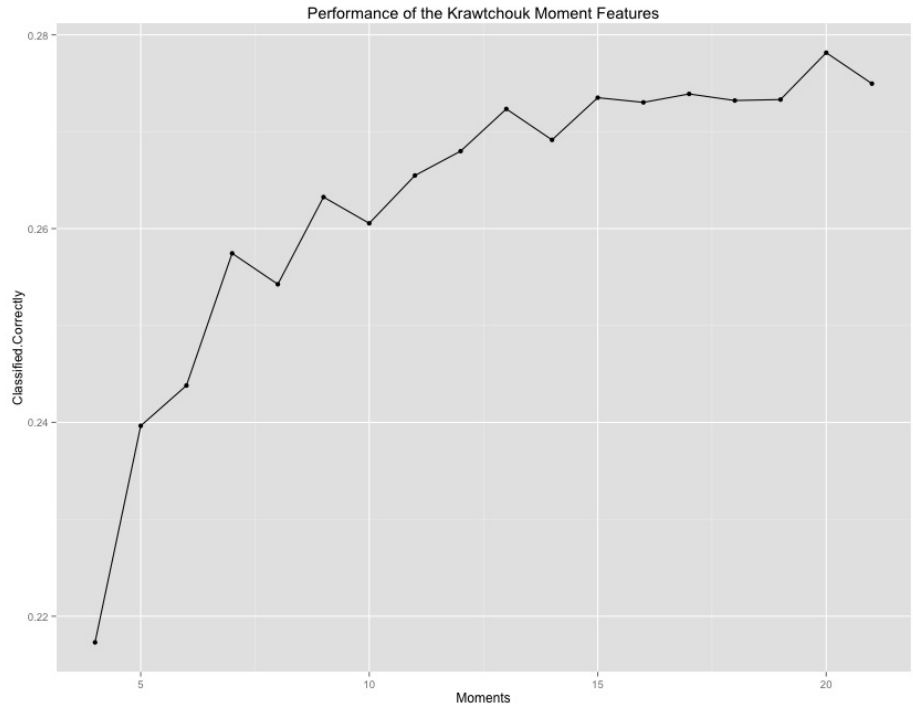
$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \bar{K}_n(x; p_1, N-1) \bar{K}_m(y; p_2, M-1) f(x, y),$$

where $f(x, y)$ is the pixel intensity and $(K_n(a; p, N)$ are the weighted Krawtchouk polynomials, and $n \in \mathbb{N}$ is order of the moment in the x- or y-direction.

- Kratuchok moments are invariant under scaling, rotation, and translation.

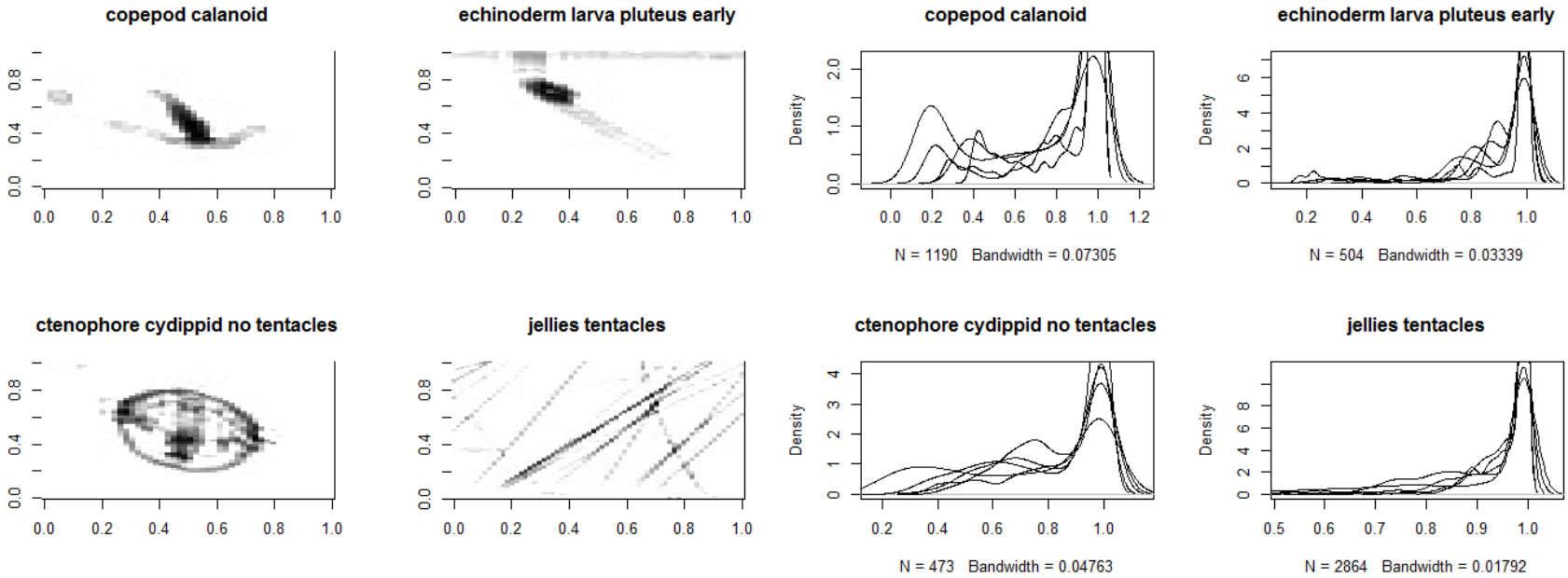
Kratuchok's Moments

- The training set was further divided into a training and validation set to determine the order of moments that yielded the "best" classification of the validation set.
- As you can see below, it appears the 20^{st} moment offers the best prediction of the validation set.



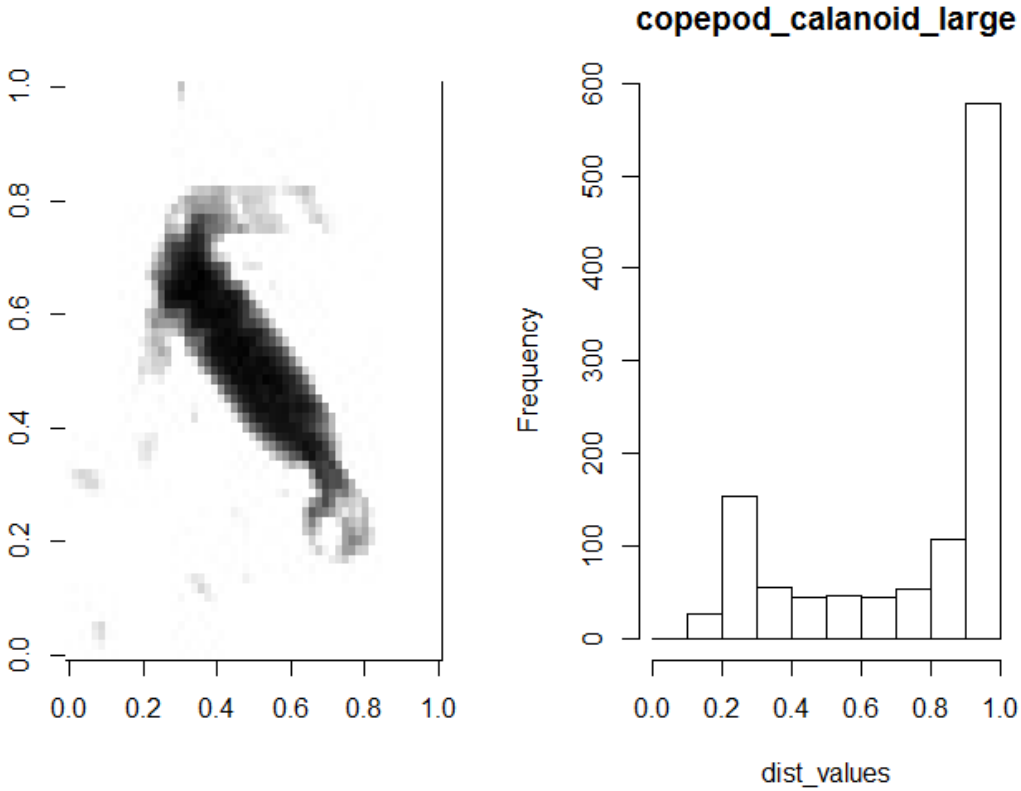
Histogram Method

- Some of species of plankton give distinct distributions of gray scale values.



Histogram Method

- The grayscale is on a [0,1] interval and we partition the interval into a width of 0.1.
- We have count the number of values that are between $[0, 0.1], [0.1, 0.2], \cdots, [0.9, 1]$.



Indicio Package and kNN

- This produces a sparse, 2048 digit feature vector for each image that can then be used to calculate the Euclidean distances between different feature vectors

Kaggle Results: The Native Touch

- 9^{th}-order Krawtchouk moments produced a score of 3.68.

713	↑3	Joe & Neema	3.677357	2	Tue, 17 Feb 2015 01:25:28 (-0h)	
-		**Black Heart**	**3.677524**	.	**Mon, 11 May 2015 07:57:38**	**Post-Deadline**

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

714	↓4	Markus Schepke	3.678106	2	Sun, 21 Dec 2014 00:01:20	

- 20^{th}-order Krawtchouk moments produced a score of 3.54.

703	↓1	noviceDM	3.542735	4	Sat, 24 Jan 2015 14:56:22	
-		**Black Heart**	**3.542917**	.	**Mon, 11 May 2015 07:53:33**	**Post-Deadline**

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

704	↓1	Mikhail Pigolkin	3.564871	4	Mon, 16 Mar 2015 23:58:33 (-0.1h)	

Kaggle Results: Histoments

- We were fairly surprise surprised as much as you.
- Histogram method produced a score of 3.29.

660	↑1	Miusay	3.288775	2	Wed, 11 Mar 2015 23:58:04	
-		Kevark	3.295601	.	Mon, 11 May 2015 09:38:29	Post-Deadline
Post-Deadline Entry If you would have submitted this entry during the competition, you would have been around here on the leaderboard.						
661	↓1	Attila Egri	3.305361	22	Mon, 16 Mar 2015 23:45:01 (-39.7d)	
662	↑1	Sinbad	3.326039	3	Wed, 21 Jan 2015 14:27:00	

- Combination of Histogram and Moments produced a score of 2.66.

597	–	Pieter Gerrit Bosma	2.653109	1	Fri, 19 Dec 2014 09:01:26	
598	–	Shicai Yang	2.657105	9	Fri, 06 Feb 2015 09:31:45 (-3.9d)	
-		Black Heart	2.668862	.	Tue, 12 May 2015 16:31:39	Post-Deadline
Post-Deadline Entry If you would have submitted this entry during the competition, you would have been around here on the leaderboard.						
599	↑2	Dan Tybor	2.668967	3	Wed, 14 Jan 2015 04:13:31	

Kaggle Results: Return to Default

- We were even more surprised with the difference in precision after setting the readJPEG argument "native" to False.
- Non-native Krawtchouk moments (20^{th}-order) produced a score of 2.18.

522	↑2	BRES 90	2.173581	6	Mon, 16 Mar 2015 23:35:47 (-0.1h)
-		**Black Heart**	**2.183565**	-	**Thu, 14 May 2015 01:21:01** Post-Deadline

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

523	↓2	39rus	2.185156	6	Tue, 23 Dec 2014 23:13:32

- Histogram and non-native Moments combo produced a score of 2.24.

528	↑3	Eyrun	2.235007	1	Fri, 27 Feb 2015 01:29:09
-		**Black Heart**	**2.239828**	-	**Thu, 14 May 2015 08:07:42** Post-Deadline

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

529	↑1	ML Enthusiasts	2.245985	14	Thu, 19 Feb 2015 03:57:38 (-19.4d)

Conclusions and Final Thoughts

- Reading images into the Raster environments significantly reduces precision.
- Krawtchouk moments required 400 features to achieve it's best Kaggle rank of 523/1049.
- The Histogram method required only 10 features to achieve it's best Kaggle rank of 661/1049.

As future work,

- Perform variable selection for dimension reduction on the Krawtchouk moment.
- Increase the number of bins measured in the Histogram method.
- Look towards 2-D filter for additional features.