

**Pitchfork** : world's most powerful hipsters

“Pitchfork is widely considered to have the power to make or break a musical career at the mere click of its mouse.”

*Brad Solomon  
Feb 2018*

**Pitchfork** : world's most powerful hipsters

“...one of the most  
colossal trainwrecks in  
indie rock history.”



**Pitchfork** : world's most powerful hipsters

“With ‘Runaway,’ he rousingly highlights his own douchebaggery, turning it into a rallying cry for all humanity. Like many of his greatest songs, it's funny, sad, and perversely relatable.”



Can we predict “Best New Music” without seeing score?

$$Y = f( \overset{X_1}{\begin{bmatrix} x_{1,1} & x_{1,1} & \cdots & x_{1,m} \\ x_{2,1} & x_{1,1} & \cdots & x_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{1,1} & \cdots & x_{m,m} \end{bmatrix}}, \overset{X_2}{\begin{bmatrix} x_{1,1} & x_{1,1} & \cdots & x_{1,m} \\ x_{2,1} & x_{1,1} & \cdots & x_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{1,1} & \cdots & x_{m,m} \end{bmatrix}} )$$

- Review content  
(TF-IDF-vectorized text)

- Review length
- Cuml. album count
- “Trend score” (author + artist)
- Reissue v. original release
- Genre

# Baseline: skewed priors

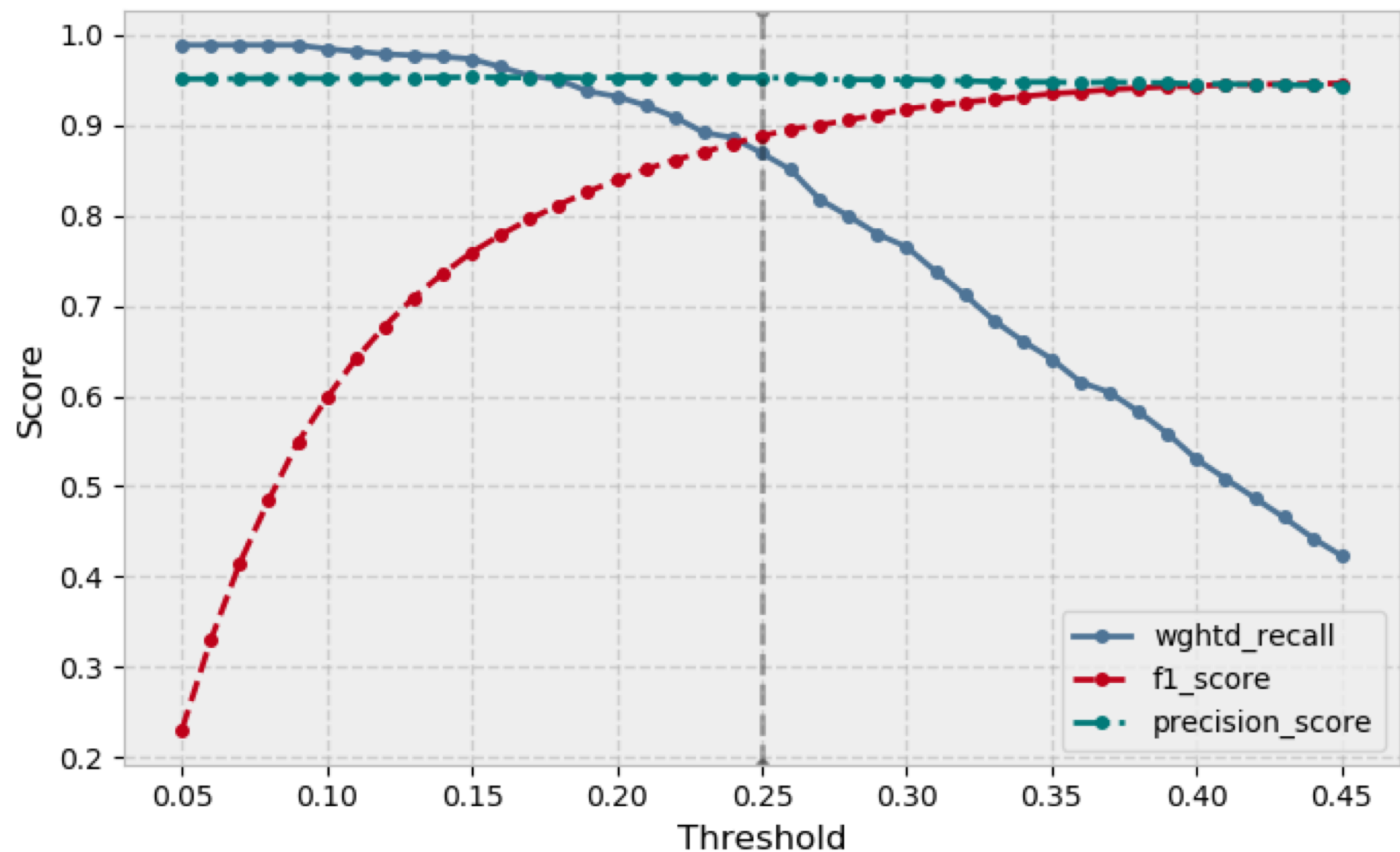
<u>Empirical frequencies</u>		
0	No label	15,038
1	Best New Music	627
2	Best New Reissue	231

Baseline accuracy score

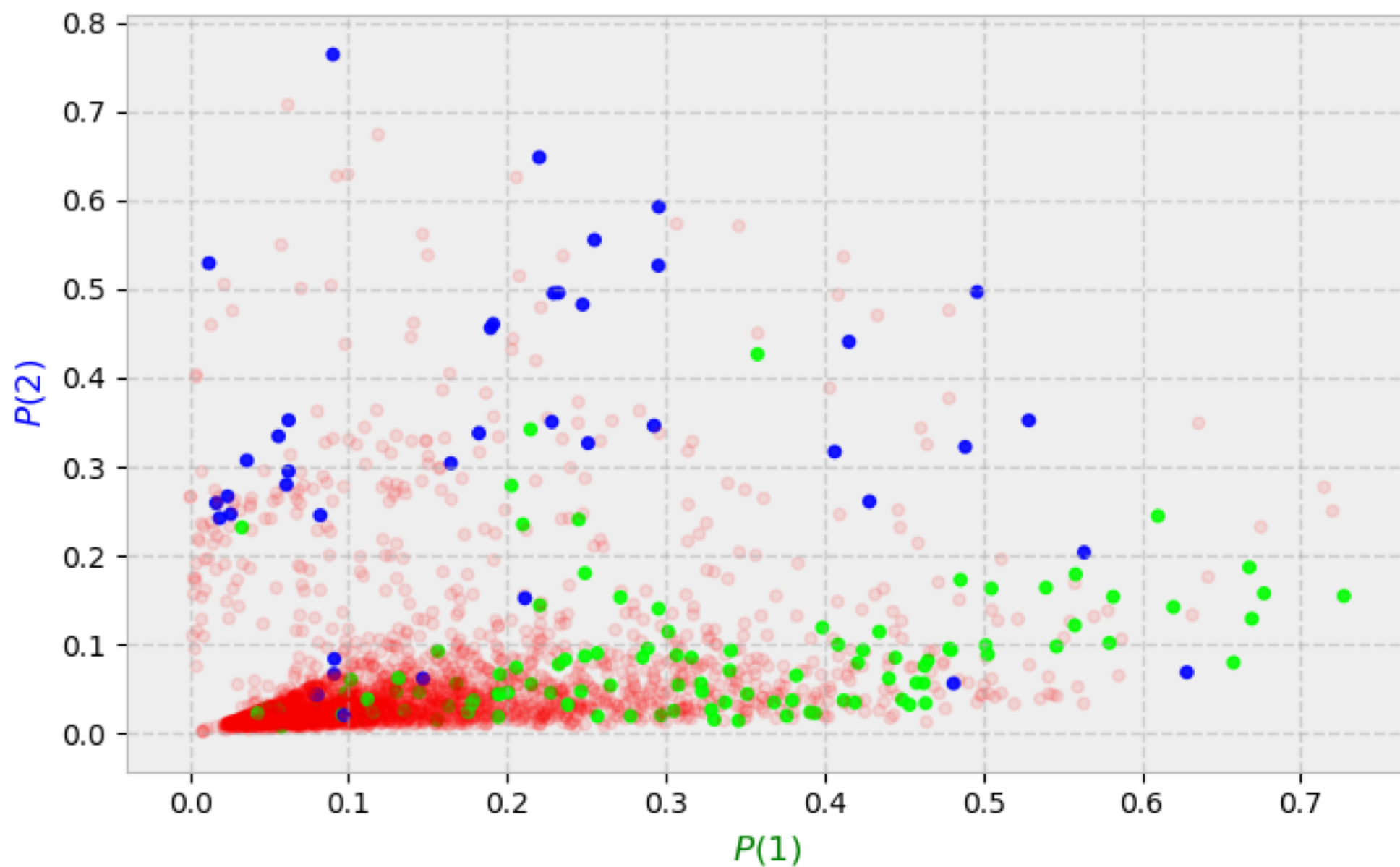
		<u>PREDICTED</u>		
		0	1	2
<u>ACTUAL</u>	0	<b>.946</b>	0.0	0.0
	1	.039	<b>0.0</b>	0.0
	2	.015	0.0	<b>0.0</b>

*Confusion matrix under (baseline) classifier  
Normalized so that entire grid sums to 1.0*

Score metrics as function of prediction threshold



Test-Set Probabilities versus Actual



# Test set classification accuracy

			<u>PREDICTED</u>		
			0	1	2
<u>ACTUAL</u>	No label	0	<b>2266</b>	364	110
	Best New Music	1	47	<b>64</b>	3
	Best New Reissue	2	10	7	<b>25</b>

		<u>PREDICTED</u>		
		0	1	2
<u>ACTUAL</u>	0	<b>0.83</b>	0.13	0.04
	1	0.41	<b>0.56</b>	0.03
	2	0.24	0.17	<b>0.60</b>

*Confusion matrix based on test set  
Normalized **row-wise***