



Andrew Ho <kironide@gmail.com>

Week 3 Day 1: Kaggle Bike Sharing Competition

2 messages

Jonah Sinick <jsinick@gmail.com>

Mon, Feb 29, 2016 at 1:18 AM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

Today we'll be looking at nonlinear models in the context of the [Kaggle Bike Sharing Competition](#). Since the optimization criterion is [Root Mean Squared Logarithmic Error](#), you'll want to model the **logarithm** of the count. Take a look at the [Kaggle Scripts](#) for some starter code.

- First try engineering many features (e.g. a dummy variable for each hour) and using regularized linear regression (glmnet). See how well you can do.

Compare your results with the results of the random forest benchmark.

- See how much better you can do than the random forest benchmark and (glmnet with engineered features) by using the methods within Chapters 7 and 8 of [Applied Predictive Modeling](#).

Jonah Sinick <jsinick@gmail.com>

Mon, Feb 29, 2016 at 1:22 AM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

See also

[A simple model for Kaggle Bike Sharing](#) by Brandon Harris.

[Quoted text hidden]