kaggle

Host   Competitions   Datasets   Kernels   Jobs   Community ▾   rbaral   Logout

Knowledge • 1,566 teams

House Prices: Advanced Regression Techniques

Tue 30 Aug 2016   Wed 1 Mar 2017 (3 months to go)

Dashboard

Home
  Data
  Make a submission

Information
  Description
  Evaluation
  Rules
  Prizes
  Timeline

Forum

Kernels
  New Script
  New Notebook

Leaderboard

My Submissions

Public Leaderboard

1. HungryFools
2. DoingItWithModels
3. andrewoohoo
4. prakhar_dwdm
5. ZavodRobotov
6. Tetsuhiro Nakane
7. Apocalypse
8. persistence
9. Attempts 2
10. Justfor

1,068 Kernels

Regularized Linear Models
131 Votes / 13 days ago / Python

Plotting all data
22 Votes / 29 days ago / R

Detailed Exploratory Data
Analysis using R
31 Votes / 47 days ago / R

XGBoost + Lasso
8 Votes / 11 days ago / Python

Ensemble Model: Stacked Model
Example
62 Votes / 2 months ago / R

Competition Details   »   Get the Data   »   Make a submission

Data Files

File Name	Available Formats
data_description	**.txt (13.06 kb)**
train	**.csv (449.88 kb)**
test	**.csv (440.83 kb)**
sample_submission	**.csv (31.19 kb)**

File descriptions

- **train.csv** - the training set
- **test.csv** - the test set
- **data_description.txt** - full description of each column, originally prepared by Dean De Cock but lightly edited to match the column names used here
- **sample_submission.csv** - a benchmark submission from a linear regression on year and month of sale, lot square footage, and number of bedrooms

Data fields

Here's a brief version of what you'll find in the data description file.

- **SalePrice** - the property's sale price in dollars. This is the target variable that you're trying to predict.
- **MSSubClass**: The building class
- **MSZoning**: The general zoning classification
- **LotFrontage**: Linear feet of street connected to property
- **LotArea**: Lot size in square feet
- **Street**: Type of road access
- **Alley**: Type of alley access
- **LotShape**: General shape of property
- **LandContour**: Flatness of the property
- **Utilities**: Type of utilities available
- **LotConfig**: Lot configuration

Lasso model for regression problem
30 Votes / 55 days ago / Python

Forum (75 topics)

Imputing missing data with the mice package in R
7 hours ago

linear regression and others
9 hours ago

Boruta Feature Importance Analysis
yesterday

Regularized Linear Models (In R)
4 days ago

Why this Linear model failed?
4 days ago

Ensemble Model: Stacked Model Example
5 days ago

teams

players

entries

- **LandSlope**: Slope of property
- **Neighborhood**: Physical locations within Ames city limits
- **Condition1**: Proximity to main road or railroad
- **Condition2**: Proximity to main road or railroad (if a second is present)
- **BldgType**: Type of dwelling
- **HouseStyle**: Style of dwelling
- **OverallQual**: Overall material and finish quality
- **OverallCond**: Overall condition rating
- **YearBuilt**: Original construction date
- **YearRemodAdd**: Remodel date
- **RoofStyle**: Type of roof
- **RoofMatl**: Roof material
- **Exterior1st**: Exterior covering on house
- **Exterior2nd**: Exterior covering on house (if more than one material)
- **MasVnrType**: Masonry veneer type
- **MasVnrArea**: Masonry veneer area in square feet
- **ExterQual**: Exterior material quality
- **ExterCond**: Present condition of the material on the exterior
- **Foundation**: Type of foundation
- **BsmtQual**: Height of the basement
- **BsmtCond**: General condition of the basement
- **BsmtExposure**: Walkout or garden level basement walls
- **BsmtFinType1**: Quality of basement finished area
- **BsmtFinSF1**: Type 1 finished square feet
- **BsmtFinType2**: Quality of second finished area (if present)
- **BsmtFinSF2**: Type 2 finished square feet
- **BsmtUnfSF**: Unfinished square feet of basement area
- **TotalBsmtSF**: Total square feet of basement area
- **Heating**: Type of heating
- **HeatingQC**: Heating quality and condition
- **CentralAir**: Central air conditioning
- **Electrical**: Electrical system
- **1stFlrSF**: First Floor square feet
- **2ndFlrSF**: Second floor square feet
- **LowQualFinSF**: Low quality finished square feet (all floors)
- **GrLivArea**: Above grade (ground) living area square feet
- **BsmtFullBath**: Basement full bathrooms
- **BsmtHalfBath**: Basement half bathrooms
- **FullBath**: Full bathrooms above grade
- **HalfBath**: Half baths above grade
- **Bedroom**: Number of bedrooms above basement level
- **Kitchen**: Number of kitchens
- **KitchenQual**: Kitchen quality
- **TotRmsAbvGrd**: Total rooms above grade (does not include bathrooms)
- **Functional**: Home functionality rating
- **Fireplaces**: Number of fireplaces

- **FireplaceQu**: Fireplace quality
- **GarageType**: Garage location
- **GarageYrBlt**: Year garage was built
- **GarageFinish**: Interior finish of the garage
- **GarageCars**: Size of garage in car capacity
- **GarageArea**: Size of garage in square feet
- **GarageQual**: Garage quality
- **GarageCond**: Garage condition
- **PavedDrive**: Paved driveway
- **WoodDeckSF**: Wood deck area in square feet
- **OpenPorchSF**: Open porch area in square feet
- **EnclosedPorch**: Enclosed porch area in square feet
- **3SsnPorch**: Three season porch area in square feet
- **ScreenPorch**: Screen porch area in square feet
- **PoolArea**: Pool area in square feet
- **PoolQC**: Pool quality
- **Fence**: Fence quality
- **MiscFeature**: Miscellaneous feature not covered in other categories
- **MiscVal**: $Value of miscellaneous feature
- **MoSold**: Month Sold
- **YrSold**: Year Sold
- **SaleType**: Type of sale
- **SaleCondition**: Condition of sale

Our Team   Careers   Terms   Privacy   Contact/Support