**Texas Tech University**

**Department of Computer Science**

**Course:** Introduction to Artificial Intelligence **Group:** 1

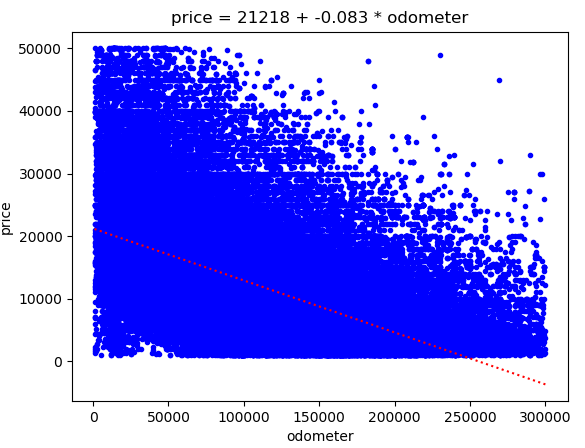
**Instructor:** Dr. Juan Carlos Rojas **Email:** [Juan-Carlos.Rojas@ttu.edu](mailto:Juan-Carlos.Rojas@ttu.edu)

**Hours:** 8:00 – 12:00 (Saturdays) **Room:** 320

# Homework 2

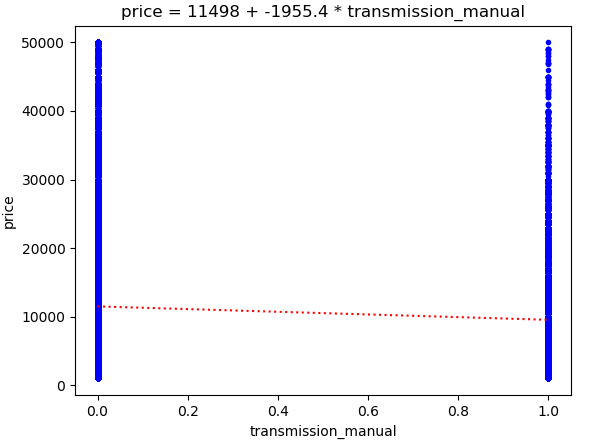
Due Saturday, June 1 at 8:00am.

## Practice 1

* Solve the Normal Equations to compute a linear-fit model for price vs. odometer of the form:   
  *y = m\*x + b*, where *x* is the vector of odometer values, and *y* is the vector of prices.
  + *m* should be a single coefficient
  + *b* should be a single value for the intersection
* Plot the scatterplot of price vs. odometer, with the best fit trendline on top
* 

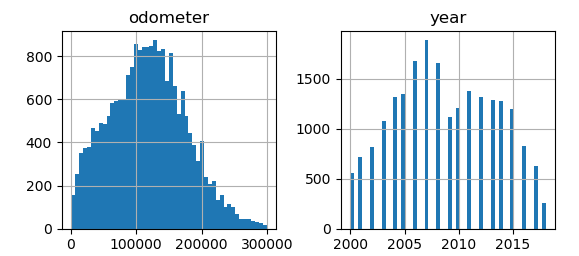
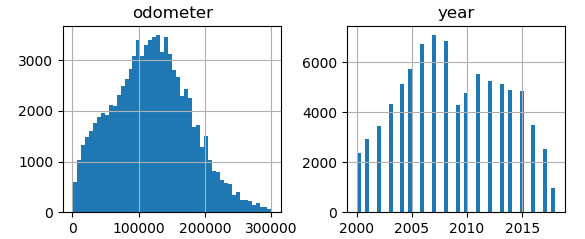
## Practice 2

* Use Pandas to label-encode and one-hot encode all categorical features in the vehicle price dataset
* Pick any one of the new dummy variables
* Compute a linear regression model of price vs. your variable
* Plot the distribution of price vs. your variable, and plot your best-fit equation on top



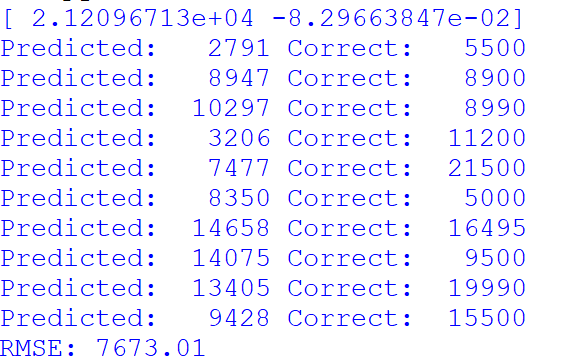
## Practice 3

* Load the vehicle price dataset pickle file (vehicle\_price\_dataset.pickle)
* Verify that the columns and rows of *train\_data* and *test\_data* are consistent with expectations
* Print the length of the *train\_labels* and *test\_labels*
  + Make sure these are also consistent
* Plot histograms of the *year* and *odometer* columns of the *train\_data* and *test\_data*
  + Make sure they are consistent with expectations



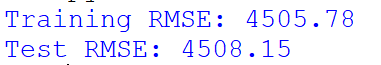
## Practice 4

* Load the vehicle price dataset pickle file (vehicle\_price\_dataset.pickle)
* Train a linear prediction model against the *odometer* feature
* Predict new labels for the test data
* Compute the RMSE of your predictions against the true labels



## Practice 5

* Load the vehicle price dataset pickle file (vehicle\_price\_dataset.pickle)
* Train a linear prediction model against **all** the available features
* Predict new labels for the test data
  + Compute the RMSE of your predictions against the true labels
* Predict new labels for the training data
  + Compute the RMSE of your predictions against the true labels



## Practice 6

* Load the credit default dataset pickle file (credit\_card\_default\_dataset.pickle)
* Train a linear prediction model against **all** the available features
* Predict new labels for the test data
  + Print the first 20 predictions, and also the first 20 correct labels

