

## Hint: Comparing the predictive power of dimension reduced data

Use the following R code hints to get started on our extension exercise.

```
Get the PCs in a data frame.
  pc_data <- h2o.predict(pca_model, covariates)</pre>
Take the first 5.
  pc_data <- pc_data[,1:5]</pre>
Add the response to the data frame.
  pc_glm_data <- h2o.cbind(pc_data, response)</pre>
Split for training and validation. Enter one line at a time.
  pc_glm_split <- h2o.splitFrame(pc_glm_data, ratios=0.75)</pre>
  pc_train_data <- pc_glm_split[[1]]</pre>
  pc_validation_data <- pc_glm_split[[2]]</pre>
Fit the predictive Model: Linerar Regression.
  glm_pca_model = h2o.glm(x = 1:5,
                       y=6,
                       training_frame = pc_train_data,
                       validation_frame = pc_validation_data,
                       max_iterations = 100,
                       solver="L_BFGS",
                       family="gaussian",
                       link="identity",
                       alpha = 0,
                       lambda = 0,
                       intercept = T)
View the model.
   summary(glm_pca_model)
```