We're going to look at how to build data pipelines, data processing systems, on Google Cloud platform.

Human and ML

* One off question: Need human insight, to collect data and analysis this particular case.
* ML: Make decision over and over again based on data

BI and ML

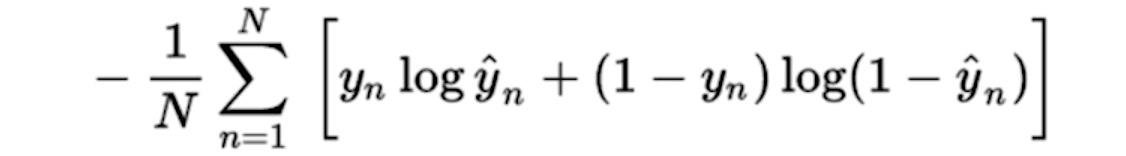
* BI: History.
* ML: Predict

ML: Pattern recognition on data

Gradient decent: Start arbitrary and try multiply time to get a good optimum(might not global optimum)

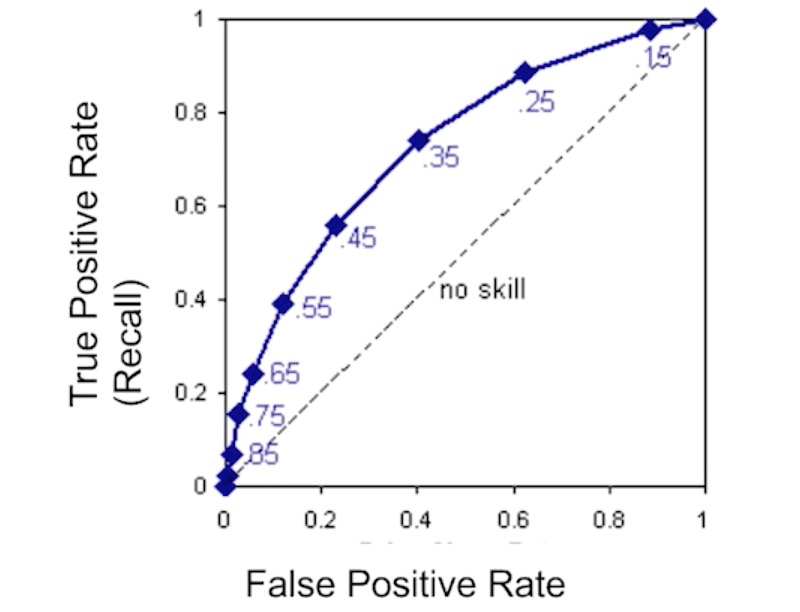
Recognize Hand write number: Softmax output probability of each digit.

Cross Entropy: Most common used error measure in classification problems



Confusion Matrix

* Accuracy: Fails when dataset unbalance. (TP+TN)/(P+N)
* Precision: TP / (TP+FP)
* Recall: 10 space, 990 full, ML recognize 1 space but get 99.1% accuracy. So we should also check recall (10%). TP/(TP+FN), TP/P
* F1
* ROC curve: X:FPR, Y: TPR, curve on threshold



AUC

**Methodology**

Pipeline

**Data**

* Data should cover all cases, enough for each
* Negative examples
* Outlier: Analyses the root cause of the outlier point => Data collect strategy. Throw away makes system not work on this particular scenario.

**Cases**

* Tips/Bill, + Gender
* Years of service / Salary:
* Pregnant weeks / age, cigarette, alcohol, weight gain
* Email SPAM filter
* New York taxi fare
* Hand write number

**Healthcare and life Sciences**

* Alerts and diagnostics from real-time patient data
* Disease identification and risk satisfaction
* Patient triage optimization
* Proactive health management
* Healthcare provider sentiment analysis



Hands on

1. Install Docker: Install Docker starting from <https://www.docker.com/products/docker>
2. Run

docker run -d -p 80:80 --name webserver nginx

If want to close and delete:

$ docker stop webserver

$ docker rm webserver

1. Install Git: <https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>
2. Clone data to local

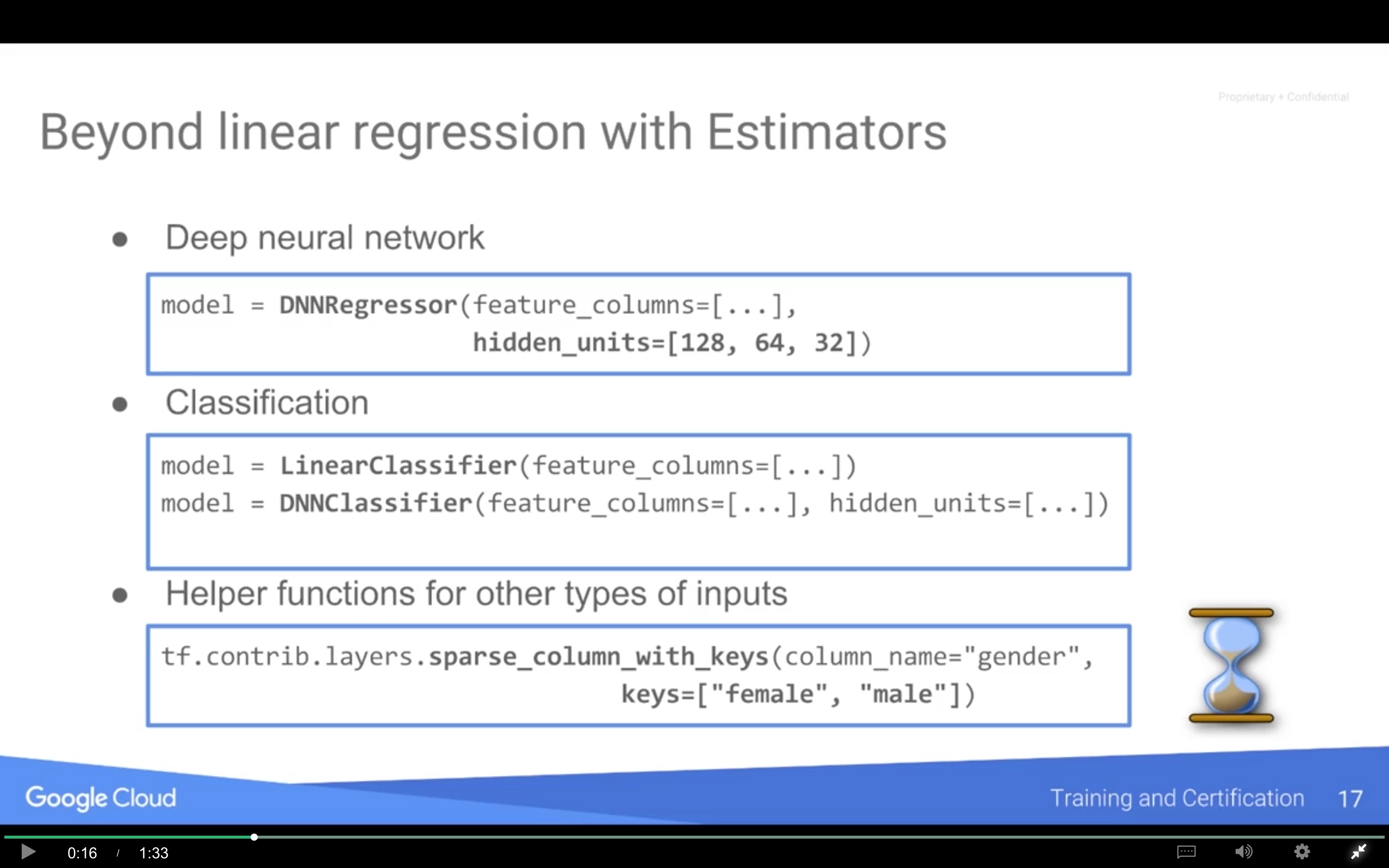
git clone https://github.com/GoogleCloudPlatform/training-data-analyst

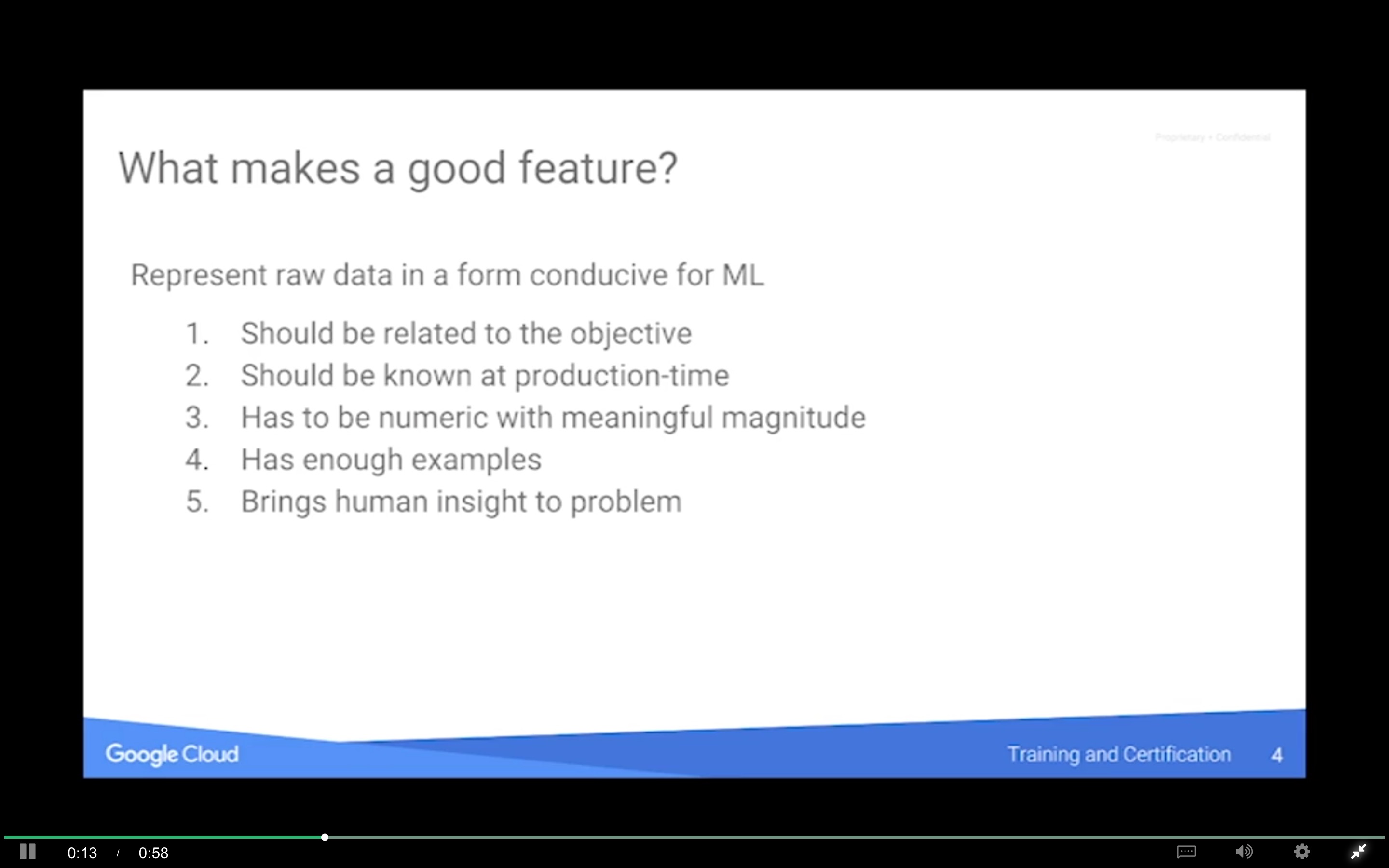
1. Start Datalab Docker container:

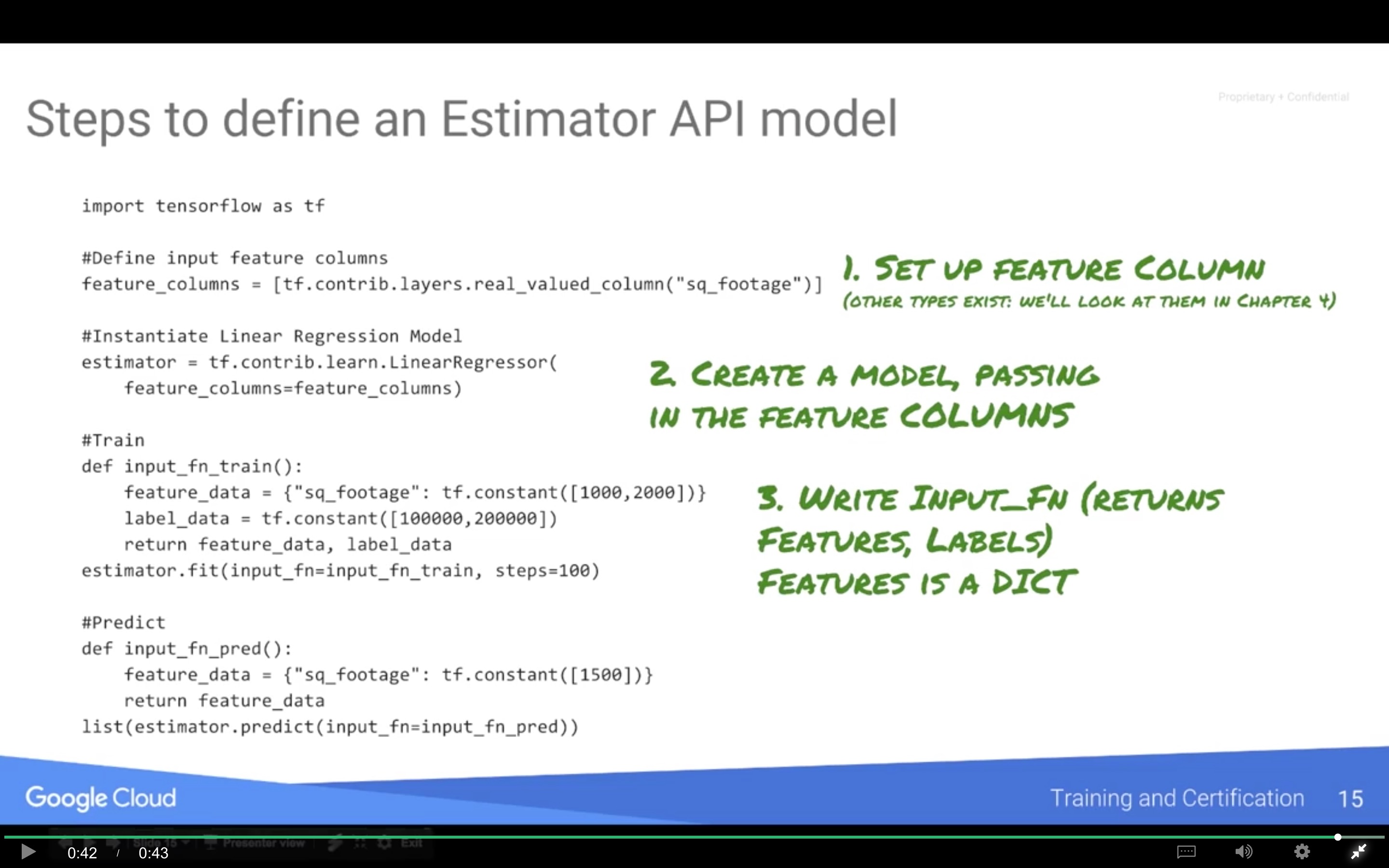
cd training-data-analyst/datalab/local

./start\_datalab.sh

1. Run http://localhost:8081/







Tuning Hyper Parameters

