

## EDUCATION

---

- **University of Southern California** Los Angeles, CA  
*Ph.D. in Neuroscience* Jul. 2011 – Aug. 2017
- **Tsinghua University** Beijing, China  
*Bachelor of Science in Physics* Jul. 2007 – Jul. 2011

## SKILLS

---

- **Theories:** Machine Learning, Natural Language Processing (NLP), Deep Learning, Scientific Computing
- **Languages:** Python, R, Matlab, SQL, Java, Mathematica, C#
- **Technologies:** Git, Shiny, MySQL, Spark, Kinect, CKAN, AWS, Linux

## WORK EXPERIENCE

---

- **Innovation Solutions** Santa Ana, CA  
*Data Scientist Intern* Summer 2015 & Spring 2016
  - **Summary:** Developed applications to record, analyze and report human movement kinematics and performance during Kinect video games and exercises.
  - **Data Acquisition:** Implemented an application *body-recorder* to acquire Kinect skeleton data and smooth it with an Unscented Kalman Filter (Matlab & C#). ~ 90% of noise and spikes are filtered out.
  - **Data Analysis:** Implemented an application to calculate human anatomical joint angles from Kinect skeleton data (Matlab & C#), allowing further analysis of human movement patterns in anatomical terms.
  - **Report and Visualization:** Designed and implemented a web application to report full kinematics performance history to users and physical therapists (R, Shiny & SQL) with intuitive graphics.
- **SixThirty Incubator** Pasadena, CA  
*Data Scientist* Jul 2015 – Aug 2016
  - **CKAN:** Set up CKAN Open Source Data Portal Platform in Linux environment; Deployed the platform onto AWS Cloud and Google Cloud, enabling SixThirty to publish and share datasets.
  - **LA Crime Data Analysis:** Analyzed “LA SHERIFF CRIMES” Open Data set with linear model to predict crime rate based on location and time of the day.

## RESEARCH PROJECTS

---

- **Customer Churn Prediction:** Developed regression, tree-based, and KNN algorithms to predict customer churn probability based on labeled data via Python and Spark. Achieved more than 90% accuracy.
- **Word Prediction:** Developed an application *wordpred* to predict the most likely following word in real time while a user is typing. Tokenized 2GB text data to train an ngram model. Deployed via Shiny framework.
- **Topic Modeling:** Grouped unlabeled textual documents and inferred latent semantic structures via Python. Extracted features by Term Frequency - Inverse Document Frequency (TF-IDF). Trained K-means and Latent Dirichlet Allocation(LDA) models. Identified latent topics and keywords of each document for clustering and calculated document similarity. Visualized results by Principal Component Analysis (PCA).
- **Human Movement Analysis::**
  - **Movement Planning:** Simulated human arm movement in state space via Optimal Control theory implemented by iterative Linear-Quadratic Regulator and Dynamic Programming in Matlab; Showed that a moderate movement velocity achieves the best accuracy.
  - **Movement Learning:** Analyzed 10+GB upper extremity kinematics data via Matlab, R, and SQL; Extracted movement patterns and variability patterns via PCA and other dimension reduction methods; Unraveled the relationship between movement patterns and movement learning.
  - **Movement Learning Prediction:** Developed dynamical State Space Model with Mixed Effects to investigate the effects of rehabilitation training, allowing customization of training schedule and prediction of future performance for each individual.