**Professor Baron’s Statistical Machine Learning Lectures**

**Spring-2021**

* **[01-19-2021](https://youtu.be/LnpSx78f5bs" \t "_blank)**

Introduction, syllabus, course plan, examples. Main concepts of Statistical Machine Learning.

[**01-22-2021**](https://youtu.be/5OVmCzKzOlY)

Statistical machine learning concepts. Flexible and inflexible methods. Prediction mean-squared error and its three components. Degrees of freedom. R lab #2.

[**01-26-2021**](https://youtu.be/Eg6hjtzgOzE) **(HYBRID)**

Review of mean-squared error of prediction. Tuning of SML algorithms. Practical example. Linear regression review: model, assumptions, the method of least squares, variance-covariance matrix, t-tests of significance.

[**01-29-2021**](https://youtu.be/IsxOtoPDddM)

Regression review continued. T-tests and Partial F-tests. Extra sum of squares. Categorical predictors, dummy variables, and interactions.

[**02-02-2021**](https://youtu.be/8khrUD-pKF4)

Regression: test of linearity; lack of fit and pure error. Classification: general approach; error classification rate; KNN algorithm.

[**02-05-2021**](https://youtu.be/xguEHkRIv-A) **(HYBRID)**

Classification methods. Classification rates. ROC curve, true positive and false positive rates. KNN algorithm.

[**02-09-2021**](https://youtu.be/WO-aCicwU-8)

Logistic Regression.

[**02-12-2021**](https://youtu.be/1ywnJXwsnP4) **(HYBRID)**

Logistic Regression. Prediction. Threshold selection by (1) tuning, (2) loss function and risks.

[**02-16-2021**](https://youtu.be/f_AS_oMqfYA)

LDA and QDA, linear discriminant analysis and quadratic discriminant analysis.

[**02-19-2021**](https://youtu.be/Vhs3bYB4RFE) **(HYBRID)**

LDA, QDA. Computation of posterior probabilities. Evaluation, comparison, and selection of classification methods.

[**02-23-2021**](https://youtu.be/lqBF8lGoubg)

Cross-validation and performance evaluation. Validation-set technique and leave-one-out cross-validation.

[**02-26-2021**](https://youtu.be/Dk4WYJji1og) **(HYBRID)**

### Leave-one-out cross-validation and K-fold cross-validation. Preparation for jackknife: bias and unbiasedness.

### [03-02-2021](https://youtu.be/1zlpWGK3pD4)

### Jackknife, a tool for bias estimation and bias reduction. Based on [these lecture notes](http://fs2.american.edu/~baron/627/Handouts/Jackknife.pdf).

### [03-05-2021](https://youtu.be/VFUWmgyNmlc) (HYBRID)

### Jackknife and Bootstrap

### [03-09-2021](https://youtu.be/NH85ntjf1bQ)

### Bootstrap - review, applications, examples. Details are in these [lecture notes](http://fs2.american.edu/baron/www/627/Handouts/Bootstrap.pdf) and [R labs](http://fs2.american.edu/baron/www/627/R/Labs/Resampling%20-%20bootstrap.htm).

### [03-12-2021](https://youtu.be/O3_C0mEm8Dw) (HYBRID)

### Jackknife and Bootstrap applications.

### [03-16-2021](https://youtu.be/tbdjdCq-MKw)

### High dimension and multicollinearity. Variance inflation factors.

### [03-19-2021](https://youtu.be/EvznwXLDY10) (HYBRID)

### Dimension reduction methods. Variable selection: exhaustive search, sequential search. Introduction to ridge regression. Bootstrap confidence interval for the efficacy of the Pfizer COVID-19 vaccine.

### [03-23-2021](https://youtu.be/RrPPhvnWpEs)

### Shrinkage methods. Ridge regression.

### [03-26-2021](https://youtu.be/-v_rONXc-bs) (HYBRID)

### Ridge regression and LASSO. Theory and application. L2 and L1 norms. LASSO as a variable selection procedure. Bayesian nature of ridge regresion and LASSO.

### [03-30-21](https://youtu.be/Z1ts2-atAWw)

### Principal components as a method of dimension reduction and elimination of collinearity - definition, implementation, construction.

### [04-02-2021](https://youtu.be/KmE83U4yCjg) (HYBRID)

### Principal components. Principal components regression (PCR). Partial least squares (PLS).

### [04-06-2021](https://youtu.be/NqSmLaoIrHI)

### Splines.

### [04-09-2021](https://youtu.be/30OoIPsSFVw) (HYBRID)

### Trees

### [04-13-2021](https://youtu.be/LYWmXtbZ3i4)

### Trees, bagging, and random forests.

### [04-16-2021](https://youtu.be/boj7huefxrQ) (HYBRID)

### Random forest tuning. Support vector machines: maximal margin classifier.

### [04-20-2021](https://youtu.be/lzz3FvensdE)

### Support vector machines. Tuning - cost and kernel. Introduction to clustering. Hierarchical clustering and dendrogram.

### 

### [04-23-2021](https://youtu.be/flxYkFS0F6w) (HYBRID)

### K-means clustering.

### Final project presentations:

### County-level analysis of wearing facemasks and related variables.

### Factors affecting the wine quality, based on a customer satisfaction survey.

### Conclusion of the course.

### Additional topics:

### Artificial neural networks.

### Google's PageRank algorithm.