AnalyticBridge

A Data Science Central Community

- Home
- Data Science Central
- Webinars
- AI
- Machine Learning
- Jobs
- Membership
- Search
- Classifieds
- Contact

Subscribe to DSC Newsletter

- All Blog Posts
- My Blog
- Add

12 Statistical and Machine Learning Methods that Every Data Scientist Should Know

- Posted by Takashi J. OZAKI on January 8, 2017 at 6:30am
- View Blog

Below is my personal list of statistical and machine learning methods that every data scientist should know in 2016.
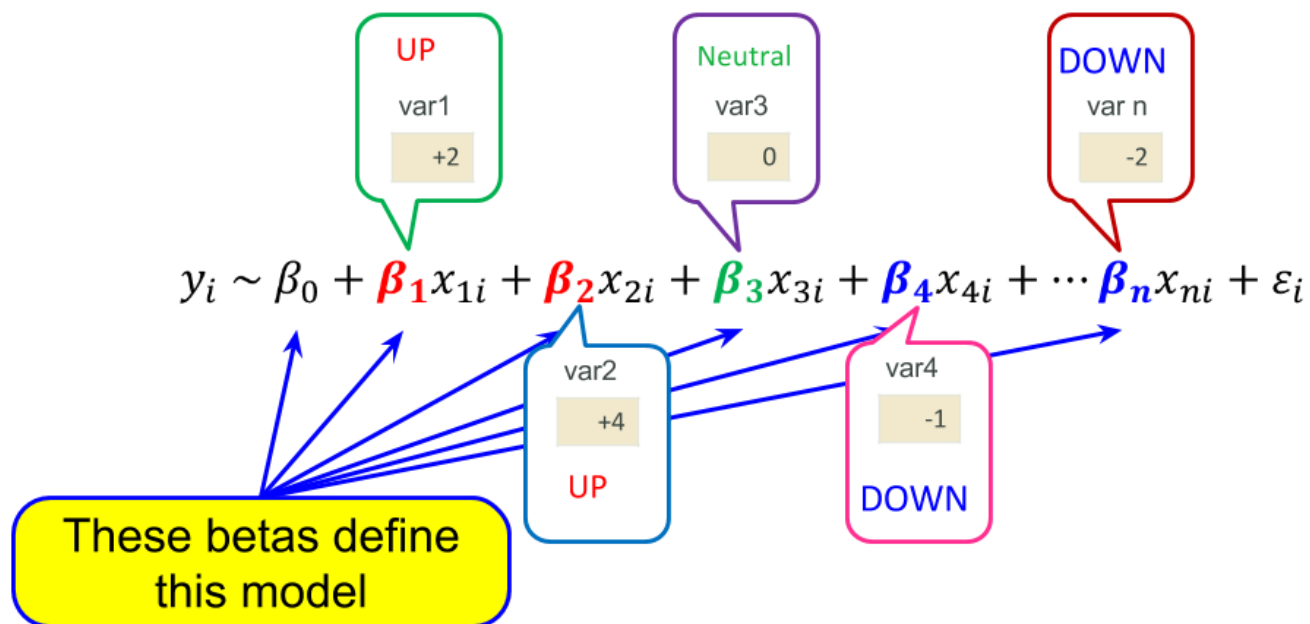
1. **Statistical Hypothesis Testing (t-test, chi-squared test & ANOVA)**
2. **Multiple Regression (Linear Models)**

3. **General Linear Models (GLM: Logistic Regression, Poisson Regression)**
4. **Random Forest**
5. **Xgboost (eXtreme Gradient Boosted Trees)**
6. **Deep Learning**
7. **Bayesian Modeling with MCMC**
8. **word2vec**
9. **K-means Clustering**
10. **Graph Theory & Network Analysis**

- **(A1) Latent Dirichlet Allocation & Topic Modeling**
- **(A2) Factorization (SVD, NMF)**

From my experience in the data science industry for 4 years, I think that currently these 12 methods are the most popular, useful and suitable for various problems requiring data science.

Betas are obtained by solving an optimization program
⇒ Magnitude and/or sign of each beta mean whether it contributes to the dependent variable or not

UP — var1 — +2
Neutral — var3 — 0
DOWN — var n — -2

$$y_i \sim \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \cdots \beta_n x_{ni} + \varepsilon_i$$

var2 — +4 — UP
var4 — -1 — DOWN

These betas define this model

As far as I've known, there have been not a few lists of "representative methods in data science" ever. However, I feel some of them are already out-of-date because they appear to neglect the latest advance of data science in the industry. Thus I made this list as the one by business person, who knows practical matters and solutions with data science, including statistics and machine learning in the industry.

In addition to the list itself, I showed R or Python scripts of an experiment on sample datasets for each method, in order to enable readers to try it easily.

The original post is **here**, including R or Python scripts and experiments on sample datasets.

Views: 32773

Like
1 member likes this

Share   Tweet   G+   Facebook

Comment

You need to be a member of AnalyticBridge to add comments!

[Join AnalyticBridge](#)

Comment by [Priya](#) on March 8, 2017 at 4:23am
Please suggest a link for budding data scientists

[RSS](#)

Welcome to
AnalyticBridge

Sign Up
or [Sign In](#)

Or sign in with:

Follow Us

On Data Science Central

- [How Big Data Helps Fight Climate Change](#)
- [Data Scientists Automated and Unemployed by 2025 - Update!](#)
- [How to measure Digital Transformation maturity?](#)
- [What is the Benefit of Modern Data Warehousing?](#)
- [In Search of Artificial General Intelligence (AGI)](#)
- [Weekly Digest, July 17](#)
- [Accelerating Convolutional Neural Networks on Raspberry Pi](#)
- [What are the Limits of Deep Learning?](#)
- [Best Data Science Books and Articles For Your Career](#)
- [Where is the value from data science?](#)
- [Deciding to adopt new technology?](#)
- [Big Data, Cognitive Extension, Self-Organizing Processes and Economic Development](#)
- [Book: Text Analytics with Python](#)
- [When Data Science Alone Won't Cut it: Deriving Signal from Observations in the Maritime Domain](#)
- [Big Data Is Changing the Game for Recruiters](#)

HEADLINES BY FEEDBURNER

Top Content

1

[Interesting Math Brain Teaser](#)

2

[50 IoT (Internet of Things) Predictions for 2016](#)

3

[Clustering idea for very large datasets](#)

4

[Career Alert, July 7](#)

- [RSS](#)
- [View All](#)

On DataViz

- [What is the future of Data visualization and Dashboard solutions?](#)
- [What Happens in 60-Seconds Online? - Infographics](#)
- [Taxonomy of 3D DataViz](#)
- [3D Data Visualisation Survey](#)
- [7 Top Data Visualization Books for Inquisitive Minds](#)
- [Why Your Brain Needs Data Visualization](#)
- [10 Dataviz Tools To Enhance Data Science](#)
- [The Top 5 Benefits of Using Data Visualization](#)
- [BI Tools for SMEs? Not Just Maybe, But DEFINITELY](#)
- [Going Deeper into Regression Analysis with Assumptions, Plots & Solutions](#)

On Hadoop

- [8 Hadoop articles that you should read](#)
- [What is Hadoop - An Easy Explanation For Absolutely Anyone](#)
- [Top 10 Commercial Hadoop Platforms](#)
- [Batch vs. Real Time Data Processing](#)
- [Is Big Data Harmful or Good?](#)
- [Lambda Architecture for Big Data Systems](#)
- [Why is there a huge buzz today around Analytics though the field has been there for decades?](#)
- [Leveraging your GPS data using Geospatial analytics](#)
- [HDFS vs. HBase : All you need to know](#)
- [An HDFS Tutorial for Data Analysts Stuck With Relational Databases](#)