

ATLAS 

Be part of the future of learning

Independent learning skills are crucial for students' success but are only mastered by a selected few

“

I am intimidated by the study material. I don't know where to start learning.

“

Insight comes from unexpected links but these links are hard to find.

“

I need to learn this skill to take the next step in my career but I am struggling to understand how to use all the free content I can find.

180 million

Learners on MOOC platforms

3%

Completion Rate

12%

Start a second course



Our mission is to give all students the tools and knowledge to master any skill by teaching them how to be a better learner.

ATLAS isn't just 'another flashcard app'

← → ↻ Not Secure | stat.yale.edu/Courses/1997-98/101/linreg.htm

Linear Regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Before attempting to fit a linear model to observed data, a modeler should first determine whether or not there is a relationship between two variables. That one variable *causes* the other (for example, higher SAT scores do not *cause* higher college grades), but that there is some significant association between two variables can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between two variables, then fitting a linear regression model to the data probably will not be helpful. A scatterplot does not indicate any increasing or decreasing trends, then fitting a linear regression model to the data probably will not be helpful. The strength of association between two variables is the [correlation coefficient](#), which is a value between -1 and 1 indicating the strength of the association.

A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The value of a is the *y* intercept (the value of y when $x = 0$).

Least-Squares Regression

The most common method for fitting a regression line is the method of least-squares. This method calculates the best-fitting line for a set of data points by minimizing the sum of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the sum of the vertical deviations is zero, the positive and negative deviations cancel each other out, resulting in a best-fitting line.

Example

Curate

Highlight important content

Generate revision questions

Organize

Store highlight history

Organize content with tags

Remember

Practice study questions

AI-driven recommendations

ATLAS isn't just 'another flashcard app'

The screenshot shows a web browser window with the URL `stat.yale.edu/Courses/1997-98/101/linreg.htm`. The page title is "Linear Regression". The main content includes a definition of linear regression, a section on "Least-Squares Regression", and an "Example" section. A yellow highlight box is overlaid on the text: "Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data." Below this, there are buttons for "Save Highlight", "Download Highlights", and "Generate Content". The "Example" section includes a scatter plot showing a positive correlation between the number of people per television set (x-axis) and the number of people per physician (y-axis). The plot shows a cluster of points at the bottom left and two outliers at the top right, with a regression line fitted to the data.

Curate

Highlight important content

Generate revision questions

Organize

Store highlight history

Organize content with tags

Remember

Practice study questions

AI-driven recommendations

ATLAS isn't just 'another flashcard app'

Unsupervised learning:
Unsupervised learning (UL) is a type of algorithm that learns patterns from untagged data. The hope is that, through mimicry, the machine is forced to build a compact internal representation of its world and then generate imaginative content. In contrast to supervised learning (SL) where data is tagged by a human, e.g. as "car" or "fish" etc, UL exhibits self-organization that captures patterns as neuronal predilections or probability densities. The other levels in the supervision spectrum are reinforcement learning where the machine is given only a numerical performance score as its guidance, and semi-supervised learning where a smaller portion of the data is tagged. Two broad methods in UL are Neural Networks and Probabilistic Methods.

Linear Regression:
In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis. Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine. Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

Machine:
A machine is any physical system with ordered structural and functional properties. It may represent human-made or naturally occurring device molecular machine that uses power to apply forces and control movement to perform an action. Machines can be driven by animals and people, by natural forces such as wind and water, and by chemical, thermal, or electrical power, and include a system of mechanisms that shape the actuator input to achieve a specific application of output forces and movement. They can also include computers and sensors that monitor performance and plan movement, often called mechanical systems. Renaissance natural philosophers identified six simple machines which were the elementary devices that put a load into motion. and calculated the ratio of

Question 1: What two main areas are in machine learning?
Answer 1: supervised learning and unsupervised learning

Question 2: What does the first term refer to?
Answer 2: prediction with human intervention

Question 3: What does unsupervised learning use?
Answer 3: historical data that has no target field

Question 4: What is the aim of unsupervised learning?
Answer 4: explore the data and find some structure or to organize it

Question 5: What is the purpose of unsupervised learning?
Answer 5: characteristics or behaviors similar to those of highly segmented marketing campaigns

Question 6: What is an example of classification?
Answer 6: spam

Question 7: What is categorized as "spam" or "legitimate"?

Curate

Highlight important content

Generate revision questions

Organize

Store highlight history

Organize content with tags





Remember

Practice study questions

AI-driven recommendations


ATLAS is the most effective way for learners to organize their content...

ATLAS 

 Brainscape  goconqr
 Repetico  Quizlet


AI-driven question generation



 Users don't lose time preparing bad questions



Custom highlighting to organize content by difficulty, topic and more



 No redundant or too difficult content revision


...and to achieve long-term retention.

ATLAS 

 Brainscape  goconqr
 Repetico  Quizlet


Fast content retrieval and personal suggestions



 Users can search all content ever collected

Active recall and spaced repetition through content linking



 First application to use unsupervised content linking through keyword extraction

Four-Stages Machine Learning Model



Keywords Extraction: Yake! and TF_IDF



Keywords Tailoring: Wikipedia

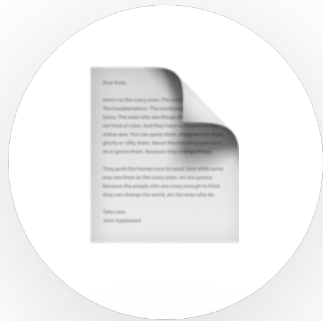


Concepts Generation: Wikipedia and Web Scraping



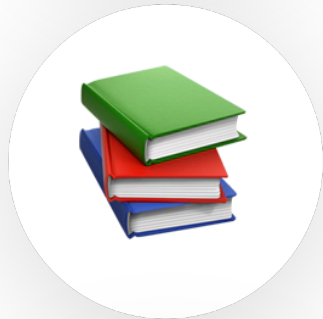
Quiz Questions Generation: T5-Transformers Based Models

Our features are powered by cutting-edge scientific research



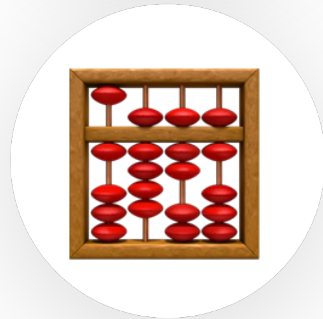
We draw on recent advancements in automatic question generation in education.

(Adapting question difficulty by highlight color; Le et al., 2014; Kurdi et al., 2020)



We integrate evidence-based study principles.

(Active recall and spaced repetition; Augustin, 2014)



We leverage state-of-the-art natural language processing algorithms.

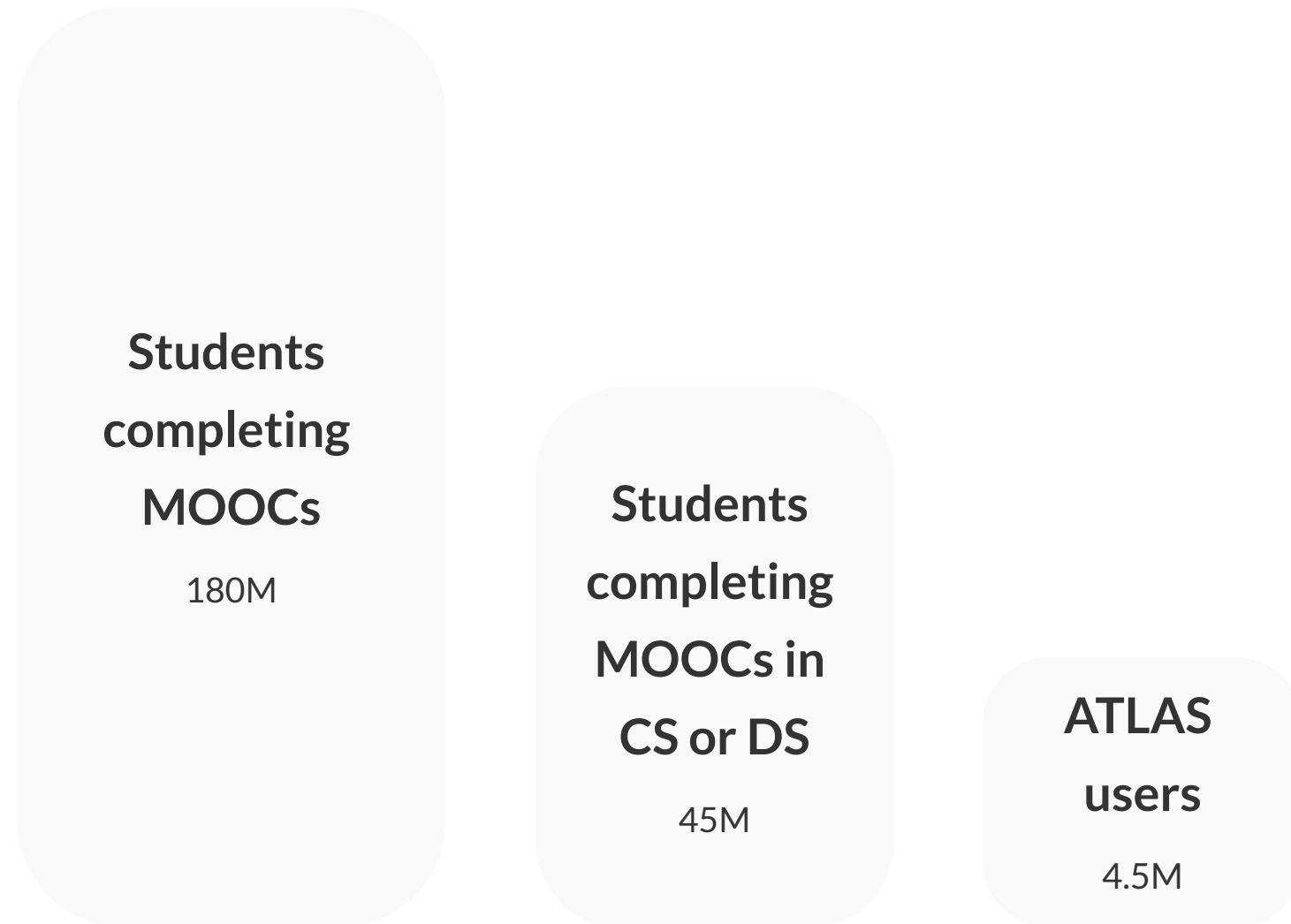
(Transformer-based end-to-end question generation; Enrico Lopez et al., 2020)

The ATLAS pricing model:
users get to love our product through the free tier and choose to go premium

\$0
Always free

\$15/m
Premium Features

We address a fast-growing market boosted by Covid-19



This analysis focuses on Computer Science and Data Science students however we plan to extend to additional subjects soon.



With a Serviceable Obtainable Market of 4.5M users and 10% Premium customers this equals a potential MRR of 6.75M.

Where are we going from here?



Launch MVP through MOOC forums, online communities, content creator and influencers



Target wider audience through referral scheme and paid marketing



Boost user growth through partnerships with MOOC providers, universities and textbook companies

We have the skill set to **fundamentally improve** how students learn



Jesse Doan
Frontend Development & AI
CS at Stanford



Conrad Borchers
Learning Science
Psychology at the University of Tübingen



Paul Muller
Business Dev & Product
CS at KCL



Mohamed Gaber
Backend Development & Data Science
Data Science at Minerva Schools at KGI

References

Augustin, M. (2014). How to learn effectively in medical school: test yourself, learn actively, and repeat in intervals. *The Yale journal of biology and medicine*, 87(2), 207.

Enrico Lopez, L., Cruz, D. K., Blaise Cruz, J. C., & Cheng, C. (2020). Transformer-based End-to-End Question Generation. arXiv e-prints, arXiv-2005.01107.

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30(1), 121-204.

Le, N. T., Kojiri, T., & Pinkwart, N. (2014). Automatic question generation for educational applications—the state of art. In *Advanced computational methods for knowledge engineering* (pp. 325-338). Springer, Cham.