



Sushrut Shendre

May 13, 2020 · 5 min read · ✨ Member-only



Search



Sushrut Shendre

79 Followers

Follow

More from Medium

Gaurav Shek... in Towards Data Scie...

Causal AI—Enabling Data Driven Decisions



Ai 12ft.io – Is the paywall gone? Yes / No

Google Data Scientist Interview Questions (Step-by-Step Solutions!)



Omolade Saliu

An End-to-End MLOps Platform Implementation using Open-source Tooling



Beverly Wang

Model Explainability: LIME & SHAP



Help Status Writers Blog Careers Privacy Terms About Knowable

12ft.io – Is the paywall gone? Yes / No

As we enter a world dictated by data and analytics, machine learning models have become the major drivers of business decisions. And as with any other business strategy, these models need to be revised with time, the technical reason behind which being ‘Model Drift’. While most course curriculums, articles, and posts define a machine learning (ML) lifecycle to start with the collection of data and to end with the deployment of the ML model in the respective environment, they forget a very important feature in the ML lifecycle, that of model drift.

What it essentially means is that the relationship between the target variable and the independent variables changes with time. Due to this drift, the model keeps becoming unstable and the predictions keep on becoming erroneous with time. Let us try to understand it from a technical viewpoint with the help of simple linear regression. In linear regression, we simply map the independent variables x_i to predict the target variable y :

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$$

where, α is the intercept, and β_i correspond to the coefficients for the variable x_i .

Often, we assume this mapping to be static, i.e. we assume that the coefficients β_i (and the intercept α) do not change with time and that the relationships governing the prediction of the target variable y will be valid for future data as well. This assumption may not hold true in all cases. And wherever it doesn't, it poses a serious threat to the business. This is because the profits of the organizations depend on such models to a great extent; and while these models might be representative of the situation at the time of development, they certainly might not hold true in the future. Owing to these changes in the underlying conditions, the predictions will start getting less accurate with time.





12ft.io – Is the paywall gone? Yes / No

Photo by [engin.akyurt](#) on [Unsplash](#)

Types of drift

Model drift can be classified into two broad categories. The first type is called 'concept drift'. This happens when the statistical properties of the target variable itself change. As is evident, if the very meaning of the variable that we are trying to predict changes, the model wouldn't really work well for this updated definition. Duh!

12ft.io – Is the paywall gone? Yes / No

The second and the more common type is 'data drift'. This happens when the statistical properties of the predictors change. Again, if the underlying variables are changing, the model is bound to fail. So obvious! A classic example of when this might happen is when the patterns in the data change due to seasonality. Whatever business model works in the summer, might not work in the winter. While flight demand surges during the holiday seasons, airlines struggle to maintain occupancy in the off-seasons. Another example is when personal preferences change, which can be related to the smartphone example mentioned in the beginning.

12ft.io – Is the paywall gone? Yes / No



12ft.io – Is the paywall gone? Yes / No

How to address this?

The best way to address this issue is to continuously re-fit the models. Based on past experiences, an estimate can be made as to when drift starts to creep in the model. Based on this, the model can be proactively re-developed so as to mitigate the risks associated with drift.

For situations where the data changes with time, weighing data can be a good option. For example, financial models deciding certain parameters based on recent transactions can incorporate features that give more weight to most

12ft.io – Is the paywall gone? Yes / No

recent transactions and lesser weights to past transactions. This not only ensures that the model is robust, but also helps keep the potential issues related to drift at bay.

A more complex methodology to fight model drift is to model the change itself. The first model developed is kept static, and serves as a baseline. Now, as a result of a change in behavior in the recent data, new models can be built to correct the predictions of this baseline model.

How often should models be re-trained?

Now that we have seen that the most common solution involves continuous re-training of the model, questions arise as to how often this needs to be done. There are multiple solutions to this, each of which differ depending on the situation.

12ft.io – Is the paywall gone? Yes / No

Sometimes, the problem will present itself. While waiting for the problem to happen is not the most elegant method, it remains the only option when it comes to new models, wherein there is no past history to understand when things might go wayward. When the problem surfaces, investigations can be made into what went wrong, and modifications can be incorporated to suppress such issues in the future.

At other times, the data related to the entities addressed in the model, observe patterns of seasonality. Here, the model should be retrained according to these seasons. For example, with an increase in spending during festive seasons, the credit lending institutes need to have special models to tackle this sudden change in patterns.

12ft.io – Is the paywall gone? Yes / No

The best way to detect drift, however, is continuous monitoring. Metrics related to the stability of the model need to be monitored at continuous intervals. This interval can be a week, a month, a quarter, etc. depending upon the domain and the business. The mode of monitoring can either be manual, or an automated script that triggers alarms and notifications whenever sudden anomalies are observed.

That brings us to the end of this article. As Heraclitus had famously said, 'Change is the only constant'. Keeping that in mind, the organizations that embrace and monitor these changes with preparedness, are the ones destined to succeed.

12ft.io – Is the paywall gone? Yes / No



Using the same
model over and
over again



Retraining the
model taking
model drift
into account

12ft.io – Is the paywall gone? Yes / No



INFO ACCOUNT

Thank you!



12ft.io – Is the paywall gone? Yes / No

More from Towards Data Science

Follow

Your home for data science. A Medium publication sharing concepts, ideas and codes.

Read more from Towards Data Science

Recommended from Medium

Rudradeb ... in Towards Data Sci...

Identifying rooftops on low-resolution images with Masked R-CNN model



Jon-Pierre Stoermer

Processing MicaSense Multispectral Imagery with DroneMapper



12ft.io – Is the paywall gone? Yes / No

James Th... in Towards Data Scie...

Random Forest for Feature Importance



Jalal Mansoori in Nerd For Tech

How I completed Machine Learning course taught by Prof Andrew Ng on Coursera ?



Monodeep Mukherjee

Papers based on Contrastive Learning



Monodeep Mukherjee

Research Papers based on Lasso Regression part2(Machine Learning)



Parsa Heidary Moghadam

Faster RCNN: a survey



Monodeep Mukherjee

How Gravitational Lensing works part2(Astrophysics 2022 edition)



12ft.io – Is the paywall gone? Yes / No