Machine Learning Challenge 1-Round 1

00D : 05H : 01M : 02S

Questions

1. Machine Learning Challenge . ()

Attempts Remaining

✅ 0 / 15

> Note:
> - You can do multiple submissions.
> - Your highest score will be considered

Machine Learning Challenge

Data Description

The dataset as provided is a summary of data available about various mutual funds and it is our task to identify which would be the most profitable funds as per basis point spread of the fund. The dataset contains details of all kinds of rated funds under a title but for this challenge we are only clustering the aaa rated funds. Please read the description below to understand the significance of each variable:

Isn't it a difficult job looking at a seed and trying to guess which one would grow up to be a tree of fruits and which one would die at the hint of hardships. Well the same is not always true for choosing your funds, owing statistical relationships the related data pose. This document is to help one cut through the jargons related to fund analysis so that it may help you choose the significant variables to work with in order to predict our target variable into clusters. Please study the following terms which would enhance your understanding of funds so that you may find relevant statistical relationships.

Help ⏾

Basis Point Spread (Value of bonds_aaa, bonds_aa….. Variable signifies the bps for the aaa shares, aa shares and so on for the same fund): The basis point is commonly used for calculating changes in interest rates, equity indices, and the yield of a fixed-income security. It is common for bonds and loans to be quoted in basis point terms. For example, it could be said that the interest rate offered by your bank is 50 basis points higher than LIBOR (London Interbank Offered Rate). A bond whose yield increases from 5% to 5.5% is said to increase by 50 basis points; or interest rates that have risen 1% are said to have increased by 100 basis points. If the Federal Reserve Board raises the target interest rate by 25 basis points, it means that rates have risen by 0.25% percentage points. If rates were at 2.50%, and the Fed raised them by 0.25%, or 25 basis points, the new interest rate would be 2.75%.

By using basis points in conversation, traders and analysts remove some of the ambiguity that can arise when talking about things in percentage moves. For example, if a financial instrument is priced at a 10% rate of interest and the rate experiences a 10% increase, it could conceivably mean that it is now 0.10 x (1 + 0.10) = 11% OR it could also mean 10% + 10% = 20%. The intent of the statement is unclear. Use of basis points in this case makes the meaning obvious: if the instrument is priced at a 10% rate of interest and experiences a 100 bp move up, it is now 11%. The 20% result would occur if there was instead a move of 1,000 bps.
(Courtesy Investopedia.com)

Save & Exit

00D :05H : 01M : 02S

Asset management companies (AMCs) employ highly qualified professionals to track developments in equity, debt and money markets and then transact accordingly in the asset markets to attain the objectives that are stated in the fund's offer documents. For such specialised services, the AMC pays a management fee to the professionals.

Read more at:
Net Expense Ratio (http://economictimes.indiatimes.com/articleshow/12410645.cms?
utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst)

Portfolio Cash Level (portfolio_cash variable): Every mutual fund scheme comes with a mandate to invest in certain type of securities. And at all times, as well. But every mutual fund scheme is allowed a tiny part of its portfolio in cash. This is allowed to meet redemptions or any 'buy' opportunities that the fund may come across on any day.
Usually, equity funds hold cash between 1% and 5% of a fund's corpus, though some funds can hold as high as 7-10% of their corpuses in cash. Some funds prefer to hold larger portions of their portfolios in cash because their mandate allows them to hold high cash levels if—as per their analysis—good stocks are not available at desirable valuations. Few others like dynamic equity funds also hold higher cash if they feel equity markets are overheated.
(Courtesy: Livemint.com)

Portfolio_Bonds,Portfolio_Stocks and Portfolio_others Variable: The percentage of total holdings held in bonds and in stocks. The sum of the three variables plus portfolio cash, preference and convertibles is 100.

Convertibles (Portfolio_convertible variable expressed in percentage): Convertibles are securities, usually bonds or preferred shares, that can be converted into common stock. Convertibles are most often associated with convertible bonds, which allow bond holders to convert their creditor position to that of an equity holder at an agreed-upon price. Other convertible securities can include notes and preferred shares, which can possess many different traits.
Read more: Convertibles (https://www.investopedia.com/terms/c/convertibles.asp#ixzz5U1IaPmEt)

Preferred Stock (Portfolio_preferred variable expressed in percentage): A preferred stock is a class of ownership in a corporation that has a higher claim on its assets and earnings than common stock. Preferred shares generally have a dividend that must be paid out before dividends to common shareholders, and the shares usually do not carry voting rights.
Preferred stock combines features of debt, in that it pays fixed dividends, and equity, in that it has the potential to appreciate in price. The details of each preferred stock depend on the issue.
Read more: Preferred Stock (https://www.investopedia.com/terms/p/preferredstock.asp#ixzz5U1JNo2dF)

Sectors variables (sectors_basic_materials,sectors_financial_services.....): The percentage of investable assets spread. It defines the percentage of asset investment on various sectors. The sum of sector variables in 100.

Year To Date Returns (returns_ytd Variable expressed in percentage): Year to date (YTD) refers to the period beginning the first day of the current calendar year or fiscal year up to the current date. YTD information is useful for analyzing business trends or comparing performance data, and the acronym often modifies concepts such as investment returns, earnings and net pay.
Read more: Year To Date (YTD) (https://www.investopedia.com/terms/y/ytd.asp#ixzz5U1Sm7diQ)

Annualized return of a year (basis) (returns_2010,returns_2011....): 'Return' is the yield that an investment generates over a period of time. It is the percentage increase or decrease in the value of the investment in that period. Returns on mutual funds are expressed in 2 different ways, viz, absolute and annualized. The most popular one being the annualized return or CAGR (Compounded Annual Growth Rate).

Help

00D : 05H : 01M : 02S

months is given in absolute basis and the returns from 1 year and above are given in absolute basis. So when you see a 5% under the 3-month column, it means the fund has given 5% in 3 months' time. 12% annualized return in 3 years means 12% return earned every year for the past three years and not 12% total return in 3 years.

Albert Einstein hasn't simply said that compound interest is the 8th wonder of the world. Compounding can do wonders to your money. 12% annualized return can double your money in 6 years. And 15% annualized return can double your money in less than 5 years! Below is the table that shows how long different interest rates take to double your money.

Absolute Return

Absolute returns, also known as point-to-point returns, calculate the simple returns on initial investment. To calculate this return all one needs is the initial and ending NAV (present NAV). In this method, the duration of holding the fund is not important. One usually uses absolute returns to calculate returns for a period of less than one year.

Formula for absolute returns

Absolute returns = ((Present NAV – Initial NAV)/ Initial NAV) *100

Annualized Return

Annualized return is the amount of money the investment has earned for the investor per annum. CAGR is compounding of returns earned over a period of time. It provides a snapshot of the of an investment's performance but doesn't give investors any indication about the volatility. Using annualized return gives a clearer picture when comparing various mutual funds that have traded over different periods of time. However, this is applicable only if you re-invest your gains every year.

Formula for annualized return

Annualized return = ((1 + Absolute Rate of Return) ^ (365/no. of days)) – 1

OR

Annualized return = ((1 + Absolute Rate of Return) ^ (1/no. of years)) – 1

Help

Alpha risk parameter for final 3 year assessment (alpha_3y variable): Alpha is a measure of an investment's performance on a risk-adjusted basis. It takes the volatility (price risk) of a security or fund portfolio and compares its risk-adjusted performance to a benchmark index. The excess return of the investment relative to the return of the benchmark index is its "alpha." Simply stated, alpha is often considered to represent the value that a portfolio manager adds or subtracts from a fund portfolio's return. An alpha of 1.0 means the fund has outperformed its benchmark index by 1%. Correspondingly, an alpha of -1.0 would indicate an underperformance of 1%. For investors, the higher the alpha the better.

Read more: Risk Parameters (https://www.investopedia.com/investing/measure-mutual-fund-risk/#ixzz5U1kaI0bk)

Beta risks parameter for final 3 year assessment (beta_3y variable):
Beta, also known as the "beta coefficient," is a measure of the volatility, or systematic risk, of a security or a portfolio in comparison to the market as a whole. Beta is calculated using regression analysis, and you can think of it as the tendency of an investment's return to respond to movements in the market. By definition, the market has a beta of 1.0. Individual security and portfolio values are measured according to how they deviate from the market.

A beta of 1.0 indicates that the investment's price will move in lock-step with the market. A beta of less than 1.0 indicates that the investment will be less volatile than the market. Correspondingly, a beta of more than 1.0 indicates that the investment's price will be more volatile than the market. For example, if a fund portfolio's beta is 1.2, it's theoretically 20% more volatile than the market.

investments.

Read more: Risk Parameters (https://www.investopedia.com/investing/measure-mutual-fund-risk/#ixzz5U1lKtSsl)

3 year mean annual return (mean_annual_return_3y variable): The average of individual percentage returns of last three years

Standard Deviation (standard_deviation_3y Variable): Standard deviation (SD)
measures the volatility the fund's returns in relation to its average.
It tells you how much the fund's return can deviate from the historical mean return of the scheme. If a fund has a 12% average rate of return and a standard deviation of 4%, its return will range from 8-16%.

Computation:
Standard Deviation (SD) = Square root of Variance (V)
Variance = (Sum of squared difference between each monthly return and its mean / number of monthly return data.

Sharpe Ratio (sharpe_ratio_3y Variable): The Sharpe ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk. Subtracting the risk-free rate from the mean return, the performance associated with risk-taking activities can be isolated. One intuition of this calculation is that a portfolio engaging in "zero risk" investment, such as the purchase of U.S. Treasury bills (for which the expected return is the risk-free rate), has a Sharpe ratio of exactly zero. Generally, the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return.

Read more: Sharpe Ratio (http://%20https://www.investopedia.com/terms/s/sharperatio.asp#ixzz5U1o0nBxG)

Treynor Ratio (treynor_ratio_3y Variable): The Treynor ratio, also known as the reward-to-volatility ratio, is a metric for determining how much excess return was generated for each unit of risk taken on by a portfolio. Excess return in this sense refers to the return earned above the return that could have been earned in a risk free investment. Although there is no true risk-free investment, treasury bills are often used to represent the risk-free return in the Treynor ratio. Risk in the Treynor ratio refers to market risk, as measured by beta. Beta measures the tendency of a portfolio's return to change in response to changes in return for the overall market.

Read more: Treynor Ratio (https://www.investopedia.com/terms/t/treynorratio.asp#ixzz5U1rFVbYG)

Objective of the problem
The objective of the problem is to predict which basis point spread zone of aaa rated mutual fund might fall into considering an investment from 2010 to 2017. The idea is to find funds with maximum yield potential and to figure out how bps varies in correlation to other fund variables.
The Basis point spread range is divided into 3 clusters which are:
1) Less than 50
2) 50 to less than 100
3) 100
The sample submission file contains the the serial number wise sorted data as per clusters, going from lower values to higher values, as in the order above. Please note that the sample submission is only for demonstration of how a the correct prediction file should be uploaded.
Please note that the training data is only for creating your data model and all predictions are to be made as per serial numbers on the test file.
- Please predict the values of bonds_aaa variable for the test dataset and sort it into a list as per clusters. The sorted file which is to be uploaded be a continuous file upto all serial numbers in the test data.

must be preserved in whichever cluster might it be filtered to.

Submission file explanation: 00D : 05H : 01M : 02S

Let's say we need to predict the value of variable x against serial numbers which are unique identifiers. Let's take an example of 6 total data points which are sorted into 3 clusters which have 2 values each.

Cluster 1: value from 5-6

Cluster 2: value from 6-7

Cluster 3: value from 7-8

Serial Number X

3 5.3

5 5.9

1 6.2

4 6.6

2 7.1

6 7.4

The above is how the predicted and sorted (as per clusters) submission file is expected to be like above. Please also note the mapping stays instead of the position of the value in the sorted list.

Evaluation Algorithm

📊 Accuracy

	Training Set	(https://cdn.skillenza.com/files/5b07ff5d-32cc-467b-a7ea-17291224d207/Yes_Bank_Train.csv)
	Test Set	(https://cdn.skillenza.com/files/958644d8-e764-4809-bfd0-e6a5abc533ca/Yes_Bank_Test_Data.csv)
	Sample Submission	(https://cdn.skillenza.com/files/b203a8cb-fa31-475b-805b-db77fa5c241e/Yes_Bank_sample_submission.csv)

Help

Upload file   Submit