Course Updates & News

MAY 13, 2014

COMPETITION SCORES POSTED

If you submitted your Kaggle ID to us and participated in the Kaggle Competition, you should now see your Competition grade on your Progress page! We are very happy with how the competition went, and we hope you enjoyed the experience.

IMPORTANT: If you do not see a grade for the competition on your progress page, but you both submitted your Kaggle ID to us and you participated in the competition, please contact us at the email address final15071x@gmail.com, letting us know what your Kaggle Numeric ID is no later than 20:00 UTC on Thursday, May 15 (two days from today).

When assigning grades for the competition, we took into account your AUC compared to the AUC of four benchmark models: the baseline AUC of 0.5, a logistic regression model without imputation, a random forest model with imputation, and the best attained AUC of one of your classmates.

Good luck on the final exam this week. Remember that no posting on the discussion forum about the final exam is allowed. Please contact us at final15071x@gmail.com if you think there are errors or if you have clarification questions regarding the final.

- The 15.071x Course Team

MAY 9, 2014

WEEK 10 AND THE FINAL EXAM

We hope you are enjoying Week 10 so far! This is the last regular week of the course before the final exam. Remember that the Week 9 Homework Assignment is due on May 12 at 23:59 UTC. The Week 10 Homework Assignment will be due on May 19 at 23:59 UTC.

The final exam will be released on May 13 at 0:00 UTC, and it will be due on May 19 at 23:59 UTC. It consists of four problems, and is similar in structure to the homework assignments in this course. The exam is intended to test your knowledge of everything we have covered in this class (Weeks 1-6 and 8-10). The final exam has the same workload estimate as the other weeks in the course - we have estimated that it will take you 8-10 hours to complete.

We want to let you know about two important policies regarding the final exam:

(1) No discussion forum posts about the final exam are allowed. We are leaving the forum open so that you can still discuss other material, but any final exam posts will be deleted and the person who made the post may have their progress erased and account disabled. We will provide you with an email address that you can use to contact us if you have clarification questions or think there are mistakes in the exam. If you have questions or concerns about the final exam, please post them before the exam is released.

(2) We have designed the Explanations differently in the final exam than the ones you are used to in the homework assignments. The Explanations will serve as hints to give you the name of the functions you should have used, or a logical explanation of the answer. They do not provide R code.

Congrats to everyone who is still taking the course, and good luck on the final!

- The 15.071x Course Team

MAY 1, 2014

WEEK 9 AND KAGGLE COMPETITION

We hope you are enjoying Week 9 so far! This week is the first of two in a different area of analytics from what we have done so far - optimization. Next week we will be discussing an extension of linear optimization, called integer optimization. Optimization is a very powerful method, and we hope you enjoy learning about it!

Don't forget that two things are due this Monday, May 5: the Week 8 Homework Assignment, and the Kaggle Competition. Don't forget to make at least one submission on Kaggle for the competition. You can make up to 5 submissions each day, so feel free to try out different methods and experiment to get your test set AUC as high as possible! Your grade on the competition should show up on your Progress report some time next week - we will send out an announcement about the competition grading when it is finalized.

- The 15.071x Course Team

APRIL 21, 2014

IMPORTANT INFORMATION ABOUT KAGGLE COMPETITION

We want to remind everyone that the deadline to submit your Numeric ID for the Kaggle competition is today - Monday, April 21 at 23:59 UTC. If you have not already done so, please visit the Competition Registration page (https://courses.edx.org/courses/MITx/15.071x/1T2014/courseware/d32b0c36ff484c228b8117257349d0e6/47697256230642bfac34bd298c0f67a9/1 in Week 6 and submit your Kaggle Numeric ID. If you do not submit your ID to us by today, you will not receive a grade for the competition. There was a small bug that was mistakenly showing some people a message that they had not submitted their ID, even though they submitted it last week. This has now been fixed. So if you have submitted your Numeric ID to us already, you should be all set.

We also want to clarify that you are allowed to use any methods or algorithms for the competition, not just the ones we have covered in class. You can take some time to learn a new algorithm that you are interested in, or just use the methods and algorithms that we have taught in class.

A note about grading: We will be assigning competition grades using quantiles of the test set AUC performance. So the highest test set AUC will be the best possible grade on the competition. The exact thresholds for different grades will be determined by the performance of the class. While your competition grade will be assigned using test set AUC, our goal is not to penalize people for random noise in test set AUC. You will definitely be rewarded for effort on the competition, and people will not be graded differently for very small differences in AUC. The reason why we are not defining exact grading cut-offs right now is that we would like to make sure that the grading is fair given the performance of the class. Also, you can still do very well in the competition if you only use the techniques you have learned in this class.

The last day of the competition is May 5, so you have two more weeks to go! However, Week 8 will be released tonight, so the course will be continuing with regular content while the competition is still open. This week, we will be focusing on data visualization, and it is our last week that teaches new techniques in R. Weeks 9 and 10 will focus on optimization, for which we will be using a spreadsheet software. More information to come!

Enjoy the Competition!

- The 15.071x Course Team

APRIL 16, 2014

WELCOME TO WEEK 7!

We hope everyone is enjoying Week 7 so far. This is the first week of the Competition, and we are excited to be working with Kaggle to make this possible. Keep in mind that this is the first time Kaggle is teaming up with edX to offer a competition like this - we are very excited about it, but hang in there with us if we encounter some bugs along the way. You are part of a very exciting learning experience! Also, keep in mind that the competition is intended to challenge you to put together everything you have learned in this class to get the best model possible on an interesting dataset.

IMPORTANT ANNOUNCEMENT ABOUT COMPETITION REGISTRATION: Since there was some confusion on the deadline to submit your Numeric ID for the Kaggle competition, we have extended the competition registration deadline to this Monday, April 21 at 23:59 UTC. If you have not already done so, please visit the Competition Registration page (https://courses.edx.org/courses/MITx/15.071x/1T2014/courseware/d32b0c36ff484c228b8117257349d0e6/47697256230642bfac34bd298c0f67a9/1 in Week 6 and submit your Kaggle Numeric ID. If you do not submit your ID to us by Monday, you will not receive a grade for the competition.

A final note about the competition: we will NOT by moderating the discussion forum on Kaggle, so if you have questions about the competition that you would like to address to staff, please do so on the edX discussion forum. However, our course Discussion Forum Guidelines (https://courses.edx.org/c4x/MITx/15.071x/asset/ForumGuidelines_15071x.pdf) apply to the Kaggle discussion forum as well, so please adhere to the guidelines in all discussion forum posts.

Remember that the Week 6 Assignment will be due this Monday, April 21 at 23:59 UTC. While the competition will be running until May 5, we will be releasing Week 8 next week. So we encourage you to get started with the competition this week if you can.

Enjoy the Competition!

- The 15.071x Course Team

APRIL 9, 2014

WELCOME TO WEEK 6!

We hope everyone is enjoying Week 6 so far. Remember that due to the extended deadlines, the Week 5 Assignment will be due this Monday, April 14 at 23:59 UTC, and the Week 6 Assignment is not due until April 21 at 23:59 UTC.

IMPORTANT ANNOUNCEMENT ABOUT COMPETITION REGISTRATION: We unfortunately discovered some problems with the original instructions for the Kaggle Competition Registration. Even if you have already given us your username, please visit the competition registration page (https://courses.edx.org/courses/MITx/15.071x/1T2014/courseware/d32b0c36ff484c228b8117257349d0e6/47697256230642bfac34bd298c0f67a9/) again and follow the instructions to provide us with your Kaggle Numeric ID. If you still have questions about the registration process, please post them here (https://courses.edx.org/courses/MITx/15.071x/1T2014/discussion/forum/i4x-MITx-15_071x-course-1T2014/threads/5345f27a21d1cc45fa000cc0) and we will respond as soon as possible. We apologize for any frustration this might be causing. We are doing this to make sure that your score on the competition correctly links to your edX account.

Next week, we will be starting the Kaggle Competition. This week is different from the other weeks of the course - there will not be any new lectures, recitation, or homework assignments released next week - just the competition. It is intended to challenge you to put together everything you have learned in this class to get the best model possible on an interesting dataset.

Lastly, there have been some questions regarding the enrollment of this course. There are currently over 25,000 students enrolled. About 3,300 students, or about 13% of the class, worked on the Quick Questions from Week 4. About 2,400 students, or about 9.5% of the class, worked on the Homework Assignment from Week 4. For a MOOC, these are great numbers, and we are excited that so many people are interested and sticking with this course. Also, thank you to everyone who is participating on the discussion forum and helping your fellow classmates. Keep up the good work!

Enjoy Week 6!

- The 15.071x Course Team

APRIL 3, 2014

WELCOME TO WEEK 5!

We hope everyone is enjoying Week 5 so far. Remember that due to the extended deadlines, the Week 4 Assignment will be due this Monday, April 7 at 23:59 UTC, and the Week 5 Assignment is not due until April 14 at 23:59 UTC.

There have been several questions about the deadline for registering for a certificate. We have decided to extend the deadline to **April 21** . So until April 21, you can upgrade your status to an Honor Code Certificate or an ID-Verified Certificate if you want.

Also, we want to remind everyone to look in the discussion forum if you are having problems. Other students may have found a solution that can help you. We are appreciative of all the students who provide insightful and helpful responses on the forum - it creates a great community for this class.

Enjoy Week 5!

- The 15.071x Course Team

MARCH 25, 2014

WELCOME TO WEEK 4, AND EXTENDED DEADLINES

We hope everyone is enjoying Week 4 so far, and great job finishing Week 3!

We have an **important annoucncement** about the due dates in this course. Due to a large number of requests to increase flexibility in the course, we have decided to extend the deadlines for the remaining homework assignments by one week each. The release dates have not changed, but you will have one extra week to complete the homework assignment. We have also decided to extend the deadline for the

competition by two extra weeks, so you will have three weeks total to work on it. For more information about the changes, please see the Course Schedule (https://courses.edx.org/courses/MITx/15.071x/1T2014/2891f8bf120945b9aa12e6601739c3e6/).

Enjoy Week 4!

- The 15.071x Course Team

MARCH 23, 2014

UPDATES ON WEEK 3

Keep up the great work! This is a challenging week, and we are impressed by your perseverance and dedication to learning the material.

There have been several issues with imputing the data - we provided the imputed data files for you for both the recitation and the loans homework problem. If your data doesn't match the data in the provided imputed files, please proceed with the files we provided. Remember to rebuild your model with the new data.

There have also been issues installing new packages - if you are running into problems, please search on the discussion forum. Several different solutions have been posted.

Great job so far!

- The 15.071x Course Team

MARCH 19, 2014

WELCOME TO WEEK 3!

We hope you enjoyed learning about linear regression last week, and that Week 3 is going well for you so far. This week introduces the method of logistic regression and the idea of classification problems. This week is one of the more mathematically challenging weeks in this class, but hopefully you will see how powerful logistic regression can be by the end of the week. We are excited about how the class is going so far - keep up the great work!

We want to remind everyone that slides and R script files are posted for each lecture and recitation. You can find the slides for an individual video at the bottom of the video, and the slides for an entire lecture (or recitation) at the bottom of the first video in the lecture (or recitation). You can find the R script file with the data set for the lecture or recitation, provided above the first video where we work in R.

IMPORTANT NOTE: You may get slightly different results in the recitation after running multiple imputation. Similarly, you may get slightly different results after running the imputation in the fourth homework problem (Predicting Loan Repayment). Unfortunately, this can happen on certain operating systems even if you set the random seed. We have provided the imputed datasets for both problems, so be sure to compare your results with the imputed dataset in the homework problem before continuing. If your results are different, please continue with the imputed data that we provide.

Enjoy Week 3!

- The 15.071x Course Team

MARCH 13, 2014

WELCOME TO WEEK 2!

We hope you enjoyed Week 1, and that Week 2 is going well for you so far. Similarly to last week, this week has two lectures, one recitation, and four homework assignments. This is the first week that teaches a specific analytics method - linear regression. This is just the beginning though! We will be covering many other powerful analytics methods in this course.

Thank you to everyone who completed the feedback form at the end of Week 1. We'll be providing a link to a feedback form for every week of the course, so that we can continue to improve the course in the future.

Lastly, just a few reminders about the logistics of the course:

1) If you have general questions about the course structure, please don't forget to take a look at the syllabus. In particular, we want to remind everyone that the pass threshold to earn a certificate in this course is 55%. While this is a challenging class, we believe that it will enable you to apply analytics to real-world applications in your career and your life.

2) If you want to keep track of a particular conversation on the discussion forum, don't forget to follow the post by clicking on the star in the upper-right corner of the initial post. Then you can see all of the posts that you are following by selecting "Posts I'm Following" in the "Show All Discussions" pull-down menu.

Enjoy Week 2!

- The 15.071x Course Team

MARCH 7, 2014

WEEK 1, VIDEO QUALITY, AND REGISTRATION

We hope Week 1 is going well for you so far! This week is intended to get you started in R, and the homework assignment should give you a lot of practice using the functions taught in the second lecture. Hopefully you can already see how analytics and R can be used in a wide variety of examples. To see a summary of future weeks in the course, please visit the new "Schedule" page, which you can find after the Syllabus tab on the top of this page.

Several students have reported problems with the video quality. Two different remedies have been suggested: (1) you can watch the videos directly on YouTube by clicking on the "YouTube" text on the bottom right corner of the video, or (2) you can download the video by clicking the download button below the video. Thank you to those who have reported these problems - we are doing what we can to fix it for future weeks. If you have other technical issues or bugs to report, please add a new post to the discussion by visiting the "Technical Issues and Bugs" discussion prompt following the "Welcome to Week 1" video.

Lastly, several people have asked about changing their registration status (Audit, Honor Code Certificate, or ID-verified Certificate). You have up to two weeks from the course start date to change your status, which for this class is up to March 18.

If you want to downgrade your status from an ID-verified certificate to an Honor Code Certificate or Audit, please contact the Billing department of edX. More information can be found here: https://www.edx.org/verified-certificate.

If you want to upgrade your status to get a certificate, please go to your edX Dashboard, and then click on "Challenge Yourself" in the Analytics Edge box. Then click on "Upgrade to Verified Track". If you want to sign up for an ID-verified certificate, just follow the instructions. If you want to sign-up for an Honor Code Certificate, click on "Why do I have to pay? What if I don't meet all the requirements?", and then scroll down to check the "Select Honor Code Certificate" box.

We hope you are enjoying the class so far, and we are looking forward to the weeks ahead!

- The 15.071x Course Team

MARCH 4, 2014

WELCOME TO 15.071X, THE ANALYTICS EDGE!

We are looking forward to sharing many exciting stories and examples of analytics with all of you. This course includes examples of analytics in a wide variety of industries, and we hope that you will learn how you can use analytics in your career and your life. One of the most important aspects of this course is that you, the student, are getting hands-on experience creating analytics models; we, the 15.071x course team, urge you to participate in the discussion forums and to use all the tools available to you while you are in the course.

We have pre-addressed some of your questions about this course in the "FAQ" document, downloadable through the "Course Handouts" panel on the right side of this page.

For an overview of how this course is organized, please download the "Syllabus," also available in the "Course Handouts" panel on the right side of this page. This document also contains grading standards. Note that edX courses use Coordinated Universal Time (UTC) for due dates and release times.

Additionally, we ask that you please review the rules and regulations of our course discussion forum. These are available in the "Discussion Forum Guidelines," which can be downloaded through a link in the "Course Handouts" panel, on the right. We want to encourage you to build a community through the discussion forum, but please adhere to the guidelines when posting on the forum.

EdX is a non-profit created by founding partners Harvard and MIT whose mission is to bring the best of higher education to students of all ages anywhere in the world, wherever there is Internet access. EdX's free online MOOCs are interactive and subjects include computer science, public health, and artificial intelligence.

(http://www.meetup.com/edX-Global-Community/)

(http://www.facebook.com/EdxOnline)

(https://twitter.com/edXOnline)

(https://plus.google.com/108235383044095082736)

(http://youtube.com/user/edxonline)

Terms of Service and Honor Code - Privacy Policy (https://www.edx.org/edx-privacy-policy)