

**שיטות בלמידת מכונה ולמידה עמוקה בניתוח מידע גיאו-מרחבי
(למידת מכונה בהנדסה גיאו-סביבתית – 016837)**

- שעתיים הרצאה
- שעתיים תרגול

הערכה:

- תרגילי בית (25%)
- מבחן (25%)
- פרויקט סיום (50%)

תקציר

עושר המידע הגיאו-אינפורמטיבי שנאסף בעשורים האחרונים מהאוויר והחלל מאפשר לחקור תהליכים גיאופיזיים וסביבתיים על פני כדור הארץ ובמערכת השמש בדיוק חסר תקדים. עם זאת, נפחי המידע הגדולים הביאו עימם אתגרים גדולים באיסוף, תיוג וניתוח המידע באופן אוטומטי וכמותני. לאחרונה, שיטות מתחום למידת המכונה העמוקה הביאו למהפכה בכריית מידע תמונתי, ניתוח סדרות-זמן וגילוי אנומליות. השימוש בשיטות אלו, בשילוב עם הנמקה כמותנית וכלים סטטיסטיים - המכונה לעיתים "מדע נתונים" (Data Science) - התבססו לאחרונה ככלים העיקריים המשמשים בניתוח מידע גיאו-אינפורמטיבי וחזותי. הקורס יסקור שיטות מבוססות למידה עמוקה (רשתות נוירונים) לניתוח מידע חזותי וגיאו-אינפורמטיבי, בדגש על שימוש מעשי בכלים פופולריים: מבחנים סטטיסטיים וערכי מובהקות (p), ניקוי מידע ועיבוד מקדים, גרסיה ו-SVM, למידה בלתי-מונחית (Autoencoder, PCA) ורשתות קונבולוציה כגון YOLO, ו-Mask-RCNN. בסיום הקורס ישלימו הסטודנטים פרויקט במחקר מבוסס-נתונים.

מבנה הקורס

הרצאה/תרגול	נושא	פירוט
הרצאה 1	הקדמה	<ul style="list-style-type: none"> • היכרות עם הסילבוס • מטרת הקורס • עקרונות מדע הנתונים • היכן משתמשים במידע גיאואינפורמטיבי? • סוגי מסדי נתונים גיאואינפורמטיביים • האתגרים בניתוח מידע גיאואינפורמטיבי • שיטות "מסורתיות" לניתוח מידע ויזואלי (Canny, Sobel, NMS). • האתגרים בניתוח מידע ויזואלי בשיטות מסורתיות. • הקדמה ללמידת מכונה ולמידה עמוק
תרגול 1	חזרה על אלגברה לינארית, הסתברות ורגרסיה	<ul style="list-style-type: none"> • כפל מטריצות • הפיכת מטריצה • הסתברות מותנית וחוק ביז • רגרסיה לינארית של פולינום ממעלה n
הרצאה 2	מחקר מונחה דאטה (Data driven analysis) וחזרה על מושגי יסוד בסטטיסטיקה	<ul style="list-style-type: none"> • מבוא לשיטה המדעית • מהי השערה מדעית? • כיצד שינתה מהפכת הדאטה את השיטה המדעית? • קורלציה וסיבתיות • משפט הגבול המרכזי • מבחנים סטטיסטיים ומובהקות • שיטת מונטה קרלו (פרמוטציות) • רגרסיה (ריבועים פחותים, אורתוגונלית, לוגיסטית, לאסו) • רגרסיה לא-לינארית

<ul style="list-style-type: none"> • מבחנים סטטיסטיים : <ul style="list-style-type: none"> ○ מבחן z ○ מבחן t ○ מבחן ks ○ מבחן וואלד • הערכת שגיאות באמצעות פרמוטציות/מונטה קרלו 	שיטות סטטיסטיות בניתוח דאטה	תרגול 2
<ul style="list-style-type: none"> • פונקציית האובדן (loss) <ul style="list-style-type: none"> ○ ריבועים פחותים ○ hinge ○ softmax ○ sigmoid • למידה מונחית ובלתי מונחית • בסיס נתוני האימון, הולידציה והבדיקה • היפר-פרמטרים • התאמת יתר והתאמת חסר (overfitting, underfitting) • קרוס ולידציה 	מושגי יסוד בלמידת מכונה	הרצאה 3
<ul style="list-style-type: none"> • חשבון דיפרנציאלי מטריציוני <ul style="list-style-type: none"> ○ גזירת וקטור לפי סקלר ○ גזירה וקטור לפי וקטור ○ גזירת מטריצה לפי וקטור ○ גזירת מטריצה לפי מטריצה 	מושגי יסוד בלמידת מכונה	תרגול 3
<ul style="list-style-type: none"> • סיווג באמצעות למידה מונחית (מסווגים לינאריים) <ul style="list-style-type: none"> ○ K-nearest neighbors ○ Support Vector Machine ○ Maximum likelihood • מציאת פונקציית hinge loss באמצעות אנליזה מטריציונית 	סיווג (קלסיפיקציה)	הרצאה 4
<ul style="list-style-type: none"> • CIFAR-10 • MNIST • CIFAR-10 לסינוג תמונות K-nearest neighbors 10 	סיווג תמונה	תרגול 4
<ul style="list-style-type: none"> • שיטת "מורד הגרדיינט" (Gradient Descent). • קצב ומומנטום הלמידה • SGD, Adam, Adagrad 	אופטימיזציה	הרצאה 5
<ul style="list-style-type: none"> • מציאת המינימום של פונקציות בשיטת מורד הגרדיינט • תכנות SGD 	אופטימיזציה	תרגול 5
<ul style="list-style-type: none"> • מוטיבציה • השראה ביולוגית • מנירון מלאכותי לרשת נוירונים • Fully connected neural networks • פונקציות אקטיבציה (ReLU, סיגמויד, softplus). • Backpropagation <ul style="list-style-type: none"> ○ חוק השרשרת הוקטורי 	מבוא לרשתות נוירונים	הרצאה 6
<ul style="list-style-type: none"> • גזירת פונקציות מרוכבות באמצעות Backpropagation 	תכנות רשת נוירונים לסינוג גידולים חקלאיים	תרגול 6

<ul style="list-style-type: none"> קלסיפיקציה של תמונות של גידולים בעזרת רשת FC ממומשת ב-Pytorch 		
<ul style="list-style-type: none"> אתחול המשקלות (אקראי, xavier) רגולריזציה: <ul style="list-style-type: none"> L1 ו-L2 Batch Normalization Dropout Data Augmentation Transfer learning 	רגולריזציה ו-overfitting	הרצאה 7
<ul style="list-style-type: none"> סקירת שיטות אוגמנטציה שימוש בספרייט imgaug לאוגמנטציה הוספת רגולריזציה L2 השוואת ביצועי המודל עם ובלי רגולריזציה 	רשת נוירונים לסיווג גידולים חקלאיים: שיפור ביצועי הרשת מהתרגול הקודם	תרגול 7
<ul style="list-style-type: none"> פעולת הקונבולוציה ריפוד (Padding) פסיעה (Stride) איגום (Pooling) ארכיטקטורות קונבולוציה מפורסמות 	רשתות קונבולוציה	הרצאה 8
<ul style="list-style-type: none"> איסוף מידע ותיווג בניית רשת ב-Pytorch זיהוי וכוונון היפרפרמטרים 	רשת קונבולוציה לסיווג גידולים חקלאיים: שיפור ביצועי הרשת מהתרגול הקודם	תרגול 8
<ul style="list-style-type: none"> R-CNN ונגזרותיה YOLO הערכת ביצועי רשת <ul style="list-style-type: none"> Precision, Recall mAP ציון F1 	רשתות נוירונים לניתוח מידע ויזואלי וגיאומטרי-מרחבי 1: זיהוי אובייקטים	הרצאה 9
<ul style="list-style-type: none"> איסוף מידע ותיווג בחירת מודל אימון המודל בחירת ביצועי הרשת היפר-פרמטרים 	זיהוי עומס תחבורתי באמצעות YOLO	תרגול 9
<ul style="list-style-type: none"> Mask R-CNN <ul style="list-style-type: none"> Backbone Feature Pyramid Network Region Proposal Network שיטות להפחתת התאמת-יתר בזיהוי אובייקטים בתמונה השוואה בין שיטות מסורתיות לכאלו המבוססות על למידת מכונה 	רשתות נוירונים לניתוח מידע ויזואלי וגיאומטרי-מרחבי 2: Instance Segmentation	הרצאה 10
<ul style="list-style-type: none"> איסוף מידע ותיווג אימון Mask R-CNN בחירת ביצועי הרשת היפר-פרמטרים השוואה לשיטות שאינן מבוססות למידת מכונה 	זיהוי דיונות ברחאן בנמיביה ומאדים מתמונות לוויין	תרגול 10
<ul style="list-style-type: none"> PCA Autoencoder Variational Autoencoder 	רשתות נוירונים לניתוח מידע ויזואלי וגיאומטרי-מרחבי 3: למידת מכונה בלתי מונחית	הרצאה 11

<ul style="list-style-type: none"> • ניתוח סדרות זמן של טמפרטורה גלובלית באמצעות PCA • ניתוח סדרות זמן של טמפרטורה גלובלית באמצעות Autoencoder • אילו אזורים מתחממים הכי מהר? 	<ul style="list-style-type: none"> • למידה עמוקה ולמידת מכונה בלתי מונחית למדידת שינוי בטמפרטורות הגלובלית 	תרגול 11
<ul style="list-style-type: none"> • אינטרפולציה ב-n מימדים • שיטות ל-upscaling • Upscaling ודקונבולוציה 	<ul style="list-style-type: none"> • רשתות נוירונים לניתוח דאטה לווייני: סופר-רזולוציה ו-upscaling 	הרצאה 12
<ul style="list-style-type: none"> • אימון מודל סופר רזולוציה למיפוי בניינים בערים מתמונות לוויין 	<ul style="list-style-type: none"> • סופר-רזולוציה 	תרגול 12
	<ul style="list-style-type: none"> • חזרה למבחן • + הנחיות להכנת פרויקט 	הרצאה 13 + תרגול 13
<ul style="list-style-type: none"> • Recurrent neural networks • טרנספורמרים • זיהוי אנומליות 	<ul style="list-style-type: none"> • ניתוח סדרות זמן וזיהוי אנומליות 	הרצאות בונוס

ספרות:

- Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep learning*. MIT press.
- Wilks, D.S., 2011. *Statistical methods in the atmospheric sciences*. Academic press.
- Camps-Valls, G., Tuia, D., Zhu, X.X. and Reichstein, M. eds., 2021. *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences*. John Wiley & Sons.

ספרות נוספת:

- Girshick, R., 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
- He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).