

فرم برنامه درسی (Course Plan)

دانشکده پزشکی - گروه آمار زیستی

تعداد واحد : ۳	نام درس : نظریه یادگیری آماری
مدت زمان ارائه درس : ۵۱ ساعت	مقطع : دکترای تخصصی (Ph.D)
پیش نیاز : ندارد	
مسئول برنامه : دکتر سعیده پوراحمد	

اهداف کلی این درس شامل موارد زیر می باشد :

- ۱) آشنایی با مفهوم یادگیری آماری و اهمیت آن
- ۲) تفاوت روشهای یادگیری با نظارت و بدون نظارت
- ۳) بررسی روشهای رگرسیونی خطی
- ۴) آشنایی با انواع روشهای طبقه بندی
- ۵) روشهای اعتباریابی متقاطع و بوت استرپ
- ۶) انواع روشهای انتخاب مدل شامل روشهای گام به گام، روشهای انقباضی و روشهای کاهش بعد
- ۷) آشنایی با داده های ابعاد بالا و لزوم کاهش بعد داده ها
- ۸) مدل های غیرخطی شامل رگرسیون چندجمله ای، رگرسیون محلی، مدل های هموارسازی و مدل های جمع پذیر تعمیم یافته
- ۹) معرفی روشهای درختی شامل درخت تصمیم، درخت رگرسیونی، درخت طبقه بندی و جنگل تصادفی
- ۱۰) روش ماشین بردار پشتیبان و ارتباط آن با روشهای معمول رگرسیونی
- ۱۱) آنالیز مولفه های اصلی شامل تئوری و کاربرد آن
- ۱۲) آشنایی با روشهای مختلف خوشه بندی

❖ هدف کلی

دانشجو در پایان دوره باید مفاهیم اصلی و روشهای مختلف یادگیری آماری را آموزش دیده و بتواند کاربرد هر روش به همراه داده های مورد نیاز آنها را شناسایی کند. سپس قادر به اعمال روشهای فرا گرفته بر داده های واقعی با نرم افزارهای R و SPSS بوده و بتواند نتایج را به درستی تفسیر و گزارش کند.

❖ اهداف اختصاصی

دانشجو باید بتواند:

- مفهوم یادگیری آماری و اهمیت فراگیری روشهای مختلف کلاسیک و نوین در تحلیل داده ها را دریابد.
- با روشهای یادگیری با و بدون نظارت آشنا شده و داده های مورد نیاز هر روش را شناسایی کند.
- با مفاهیم اولیه مدلسازی شامل شناسایی نوع ارتباط بین متغیرها، لزوم برقراری تعادل بین قدرت مدل در پیش بینی و تفسیرپذیری آن، روشهای ارزیابی صحت مدل و اهمیت تعادل بین واریانس و خطا در پیش بینی ها آشنا شود.
- مدل رگرسیون خطی ساده و چندگانه را با جزئیات فرا گرفته و تفاوتها و شباهتهای آن را با روش نزدیکترین همسایگی بیان کند.
- انواع روشهای طبقه بندی شامل روش رگرسیون لجستیک، تحلیل ممیزی خطی و درجه دوم و روش کازدیکترین همسایگی را فرا گرفته و نتایج هر کدام را تفسیر کند.
- روشهای بر پایه ی نمونه گیری های مجدد شامل روش اعتباریابی متقاطع و روش بوت استرپ را آموزش دیده و به کار بندد.
- انواع روشهای انتخاب مدل شامل روشهای گام به گام، روشهای انقباضی و روشهای کاهش بعد را بشناسد.
- مفهوم داده های با ابعاد بالا را دریافته و لزوم کاهش بعد داده ها برای تحلیل بهتر را درک کند.
- با انواع مدل های غیرخطی شامل رگرسیون چندجمله ای، رگرسیون محلی، مدل های هموارسازی و مدل های جمع پذیر تعمیم یافته آشنا شود.
- روشهای درختی شامل درخت تصمیم، درخت رگرسیونی، درخت طبقه بندی و جنگل تصادفی را آموزش دیده و بر داده های واقعی اجرا کند.

- با مفهوم بردار پشتیبان و مورد استفاده ی روش ماشین بردار پشتیبان آشنا شده و ارتباط آن با روشهای معمول رگرسیونی مثل رگرسیون لجستیک را فراگیرد.
- تئوری و کاربرد آنالیز مولفه های اصلی را به عنوان یک روش یادگیری بدون نظارت آموزش ببیند.
- با روشهای مختلف خوشه بندی به عنوان انواع دیگری از روشهای بدون نظارت آشنا شده و آنها را به کار گیرد.
- کدهای برنامه نویسی تمامی روشهای آموزش دیده در نرم افزار R را فراگیرد و برای تحلیل انواع داده های واقعی استفاده و نتایج را به صورت علمی تفسیر کند.

❖ هدف کلی

آشنایی با مفهوم یادگیری آماری و اهمیت آن

❖ اهداف اختصاصی

دانشجو باید بتواند:

- منظور از روشهای یادگیری آماری را بیان کند.
- مفهوم رابطه ی تابعی بین متغیرها را دریابد.
- نحوه ی برآورد تابع رابطه ی بین متغیرها را فرا گیرد.
- اهمیت برقراری تعادل بین صحت پیش بینی و تفسیرپذیری مدل را درک کند.
- معیار اندازه گیری مناسبت مدلها را فراگیرد.
- اهمیت برقراری تعادل بین واریانس و تورش (بایس) را درک کند.

❖ هدف کلی

تفاوت روشهای یادگیری با نظارت و بدون نظارت

❖ اهداف اختصاصی

دانشجو باید بتواند:

- مفهوم یادگیری با نظارت و بدون نظارت را دریابد.
- داده های مورد نیاز هر روش را شناسایی کند.
- انواع روشهای یادگیری با نظارت و بدون نظارت را بیان کند.

❖ هدف کلی

بررسی روشهای رگرسیونی خطی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- اهداف روشهای رگرسیونی را بشناسد.
- جزئیات تئوری روشهای رگرسیونی خطی را آموزش ببیند.
- منظور از هم خطی بین متغیرها را بداند.
- رگرسیون خطی را با روش نزدیک ترین همسایگی مقایسه کند.

❖ هدف کلی

بررسی انواع روشهای طبقه بندی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- اهداف روشهای طبقه بندی و تفاوت آنها با روشهای رگرسیونی را دریابد.
- روش نزدیک ترین همسایگی را در پیش بینی یک پاسخ کیفی آموزش ببیند.
- رگرسیون لجستیک را به عنوان یک روش طبقه بندی بشناسد و جزئیات تئوری آن را فرا گیرد.
- تحلیل افتراقی خطی و مورد استفاده ی آن را بداند.
- تحلیل افتراقی درجه دوم و تفاوت موارد استفاده ی آن با روش خطی را دریابد.
- انواع روشهای طبقه بندی را با هم مقایسه کند.

❖ هدف کلی

روشهای اعتباریابی متقاطع و بوت استرپ

❖ اهداف اختصاصی

دانشجو باید بتواند:

- منطق حاکم بر روشهای اعتباریابی متقاطع را دریابد.
- انواع روشهای اعتباریابی متقاطع شامل روش مجموعه اعتباریابی، اعتباریابی متقاطع Leave-one-out و k-fold را بشناسد.
- با روش بوت استرپ در ارزیابی مدل آشنا شود.
- روش اعتباریابی متقاطع و بوت استرپ را با هم مقایسه کند.

❖ هدف کلی

انواع روشهای انتخاب مدل شامل روش بهترین زیر مجموعه، روشهای گام به گام و روشهای انقباضی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- مفهوم انتخاب مدل خطی و منظم سازی آن را درک کند.
- روش انتخاب بهترین زیرمجموعه را فرا گیرد.
- روشهای گام به گام شامل روش رو به جلو، روش رو به عقب و روشهای ترکیبی را آموزش ببیند.
- معیارهای انتخاب مدل بهینه را بشناسد و آنها را به کار بگیرد.
- روشهای انقباضی در انتخاب مدل را فراگیرد.
- انواع روشهای انقباضی شامل رگرسیون رنج و لاسو را آموزش دیده و آنها را با هم مقایسه کند.

❖ هدف کلی

آشنایی با داده های ابعاد بالا و لزوم کاهش بعد داده ها

❖ اهداف اختصاصی

دانشجو باید بتواند:

- با ماهیت داده هایی با ابعاد بالا آشنا شود.
- ملاحظات موجود در مواجهه با داده های با بعد بالا را فرا گیرد.
- انجام رگرسیون در داده های با بعد بالا را آموزش ببیند.
- نتایج تحلیل داده های با بعد بالا را تفسیر کند.
- با آنالیز مولفه های اصلی و حداقل مربعات جزئی به عنوان روشهای کاهش بعد آشنا شود.

❖ هدف کلی

مدلهای غیرخطی شامل رگرسیون چندجمله‌ای، رگرسیون محلی، مدل‌های هموارسازی و مدل‌های جمع پذیر تعمیم یافته

❖ اهداف اختصاصی

دانشجو باید بتواند:

- با روش رگرسیون چند جمله‌ای و کاربرد آن آشنا شود.
- رگرسیون اسپلاین را بشناسد و به کار بندد.
- با روش هموارسازی اسپلاین‌ها آشنا شود.
- رگرسیون محلی و مورد استفاده‌ی آن را آموزش ببیند.
- مدل‌های جمع پذیر تعمیم یافته و کاربرد آنها را بشناسد.

❖ هدف کلی

معرفی روشهای درختی شامل درخت تصمیم، درخت رگرسیونی، درخت طبقه بندی و جنگل تصادفی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- مورد استفاده‌ی روشهای درختی و ماهیت آنها را دریابد.
- با روش درخت تصمیم آشنا شود.
- درختهای رگرسیونی و جزئیات تئوری آنها را آموزش ببیند.
- روشهای هرس کردن درخت را فرا گیرد.
- درخت‌های طبقه بندی و مورد استفاده‌ی آنها را بشناسد.
- روشهای Boosting, Bagging و جنگل تصادفی را آموزش ببیند.

❖ هدف کلی

روش ماشین بردار پشتیبان و ارتباط آن با روشهای معمول رگرسیونی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- مفرد استفاده‌ی روش ماشین بردار پشتیبان و مفهوم بردار پشتیبان را درک کند.
- روش طبقه بندی گر ماکزیمم حاشیه و جزئیات تئوری آن را فرا گیرد.
- روش طبقه بندی گر بردار پشتیبان را آموزش دیده و با روش ماکزیمم حاشیه مقایسه کند.
- روش طبقه بندی گر بردار پشتیبان را برای بیش از دو کلاس به کار بندد.
- این روشها را با رگرسیون لجستیک مقایسه کند.

❖ هدف کلی

آنالیز مولفه‌های اصلی شامل تئوری و کاربرد آن

❖ اهداف اختصاصی

دانشجو باید بتواند:

- با روش آنالیز مولفه‌های اصلی به عنوان یک روش یادگیری بدون نظارت آشنا شود.
- جزئیات تئوری این روش را فرا گرفته و آن را روی یک مجموعه داده به کار ببندد.
- نتایج حاصل از آنالیز مولفه‌های اصلی را تفسیر کند.
- انواع کاربردهای روش آنالیز مولفه‌های اصلی را بداند.

❖ هدف کلی

آشنایی با روشهای مختلف خوشه بندی

❖ اهداف اختصاصی

دانشجو باید بتواند:

- با روشهای تحلیل خوشه بندی به عنوان یکی از روشهای یادگیری بدون نظارت آشنا شود.
- انواع روشهای خوشه بندی شامل روش خوشه بندی کا میانگین و سلسله مراتبی را فرا گرفته و آنها را با هم مقایسه کند.
- نتایج خوشه بندی را تفسیر و گزارش کند.

روش آموزش

- آموزش تئوری: ارائه مطالب توسط استاد به صورت سخنرانی
- ارائه مطالب تدریس شده با نرم افزار توسط دانشجو
- ارائه پروژه در مباحث تخصصی توسط دانشجو
- ارائه مقالات کاربردی مرتبط با موضوعات تدریس شده توسط دانشجو

شرایط اجراء

❖ امکانات آموزشی بخش

- اسلاید پروژکتور ، ویدئو پروژکتور و کامپیوتر و نرم افزارهای آماری

❖ آموزش دهنده

- دکتر سعیده پوراحمد دانشیار گروه آمار زیستی

منابع اصلی درسی

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, 2014.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Second Edition, Springer, 2009.

ارزشیابی

❖ نحوه ارزشیابی

- ارائه مثالهای کاربردی کتاب و انجام تکالیف نرم افزاری توسط دانشجو
- ارائه مقالات کاربردی در مباحث مختلف یادگیری آماری
- امتحان میان ترم
- امتحان پایان ترم

❖ نحوه محاسبه نمره کل

- انجام تکلیف و پروژه ۳۰٪
- امتحان میان ترم ۳۰٪
- امتحان پایان ترم ۴۰٪

❖ مقررات

- حداقل نمره قبولی: ۱۴
- تعداد دفعات مجاز غیبت در کلاس: ۴ جلسه

جدول زمانبندی درس تحلیل چند متغیره کاربردی

روش ارزشیابی	امکانات مورد نیاز	منابع درسی	نحوه ارائه	ساعت ارائه	سرفصل مطالب
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	آشنایی با مفاهیم پایه ی یادگیری آماری
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	اهمیت برقراری تعادل بین صحت و تفسیرپذیری مدل و معیارهای سنجش برازش خوب
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	مدلهای رگرسیون خطی (۱)
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	مدلهای رگرسیون خطی (۲)
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای طبقه بندی و مقایسه آنها با روشهای رگرسیونی
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	انواع روشهای طبقه بندی رگرسیون لجستیک
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	انواع روشهای طبقه بندی

			کلاس		تحلیل افتراقی خطی و درجه دوم
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای اعتبار یابی متقاطع
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش بوت استرپ
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	منظم سازی و انتخاب مدل روش بهترین زیر مجموعه
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای گام به گام
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای انقباضی شامل رگرسیون ریبج و لاسو
					امتحان میان ترم
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	آشنایی با داده های با بعد بالا و ملاحظات تحلیل و تفسیر نتایج آنها
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای کاهش بعد داده ها
پرسش و پاسخ		نرم افزارهای آماری	ارائه دانشجو با نرم افزار		کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	مدلهای غیر خطی و رگرسیون چند جمله ای
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	رگرسیون اسپلاین و هموار سازی آن
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	رگرسیون محلی
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	مدلهای جمع پذیر تعمیم یافته
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای درختی و درخت تصمیم
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای درخت رگرسیونی و درخت طبقه بندی
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روشهای Bagging و Boosting و جنگل تصادفی
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R

پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش طبقه بندی گر ماکزیمم حاشیه به عنوان پایه ی روشهای ماشین بردار پشتیبان
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش طبقه بندی گر بردار پشتیبان و مقایسه آن با روش ماکزیمم حاشیه
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش طبقه بندی گر بردار پشتیبان برای بیش از دو کلاس
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	آنالیز مولفه های اصلی به عنوان یکی از روشهای یادگیری بدون نظارت
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش تحلیل خوشه بندی کا میانگین
پرسش و پاسخ	کامپیوتر	کتابهای درسی	سخنرانی و بحث در کلاس	۲	روش تحلیل خوشه بندی سلسله مراتبی
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	کارگاه عملی و حل تمرین با نرم افزار R
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	فعالیت دانشجویی (ارائه مقاله یا پروژه)
پرسش و پاسخ	کامپیوتر	نرم افزارهای آماری	ارائه دانشجو با نرم افزار	۲	فعالیت دانشجویی (ارائه مقاله یا پروژه)
					پایان ترم