بازشناسی آماری الگو ها
جلسه دوم

مفاهیم اولیه در بازشناسی آماری

مقدمه ای بر تئوری احتمال

فهرست

- یادآوری رویکرد های اساسی در PR
- تشریح ایده ها و مفاهیم اصلی طبقه بندی آماری
- تشریح مسائل اصلی در SPR
- مقدمه ای در مفاهیم اولیه احتمالات
- معرفی احتمال شرطی و فرمول Bayes
- توابع چگالی توزیع و پارامتر های معمول
- معرفی تابع توزیع نرمال

صورت بندی ریاضی مسئله در بازشناسی آماری الگو ها

X--- all the observables using existing sensors and instruments

x --- is a set of features selected from components of X or linear/non-linear functions of X.

p(w|**x**) --- is our belief/perception about the subject class with uncertainty represented by probability.

α --- is the action that we take for x.

We denote the three spaces by

$$\mathbf{x} \in \Omega^{d}, \quad \mathrm{w} \in \Omega^{C}, \quad \alpha \in \Omega^{\alpha}$$

$\mathbf{x} = (x_1, x_2, \ldots, x_d)$ is a vector

w is the index of classes, $\Omega^{C} = \{w_1, w_2, \ldots, w_k\}$

یادآوری
رویکرد های اساسی در PR

- Template Matching
  - بررسی تطابق در مورد نمونه هایی که اعضای یک کلاس تطابق کامل دارند (اسکناس ها یا فونت های نوشتار ماشینی)
- Statistical
  - هر نمونه، یک نقطه (منسجم)در فضای مشخصه هاست. عموما خصوصیات آماری مورد استفاده
- Structural
  - هر نمونه از اجزایی تشکیل شده که مستقل از هم قابل بررسی هستند (مانند کلمات که از حروف تشکیل شده اند)
- Neural Networks
  - ادوات ریاضی با رویکرد جعبه سیاه که بر اساس نگاشت ورودی- خروجی و بدون تصریح در چگونگی دستکاری و تفسیر مشخصه ها به طبقه بندی می پردازند.

یاد آوری
ایده کلی در SPR

- دیدیم که در رویکرد آماری، بازنمایی هر نمونه توسط یک نقطه (بردار در فضای d بعدی) صورت می گیرد.
- هر بعد این بردار یا مستقیما از اندازه گیری ها ناشی شده و یا طی یک پروسه استخراج مشخصه ها، غیر مستقیم به سنجش ها مربوط است.
- از آنجا که:
  - سنجش ها همیشه با مقداری خطا روبرو هستند،
  - هر مشخصه در هر یک از کلاس ها، دارای بازه ای مجاز از تغییرات است(به جای یک مقدار ضربه ای)،
- لذا
  - نقاط بازنمایی کننده نمونه ها، فی الواقع متغیر بردار های اتفاقی هستند.
  - عضویت هر نمونه در کلاس ها نیز باید با تابع احتمال بیان گردد
- کلیه اطلاعات مربوط به مسئله در غالب توابع احتمال (توزیع احتمال) بیان می شوند.

یادآوری
مسائل اصلی در SPR

- مسئله مدل بازنمایی Representation نمونه ها (بردار های اتفاقی)
  - استخراج و انتخاب مشخصه ها بر اساس پارامتر های آماری صورت می گیرد
- مسئله استخراج دانش مر بوط به مسئله(تخمین چگالی احتمال از روی نمونه ها)
  - انواع یادگیری (با مربی و بدون مربی)
  - پارامتری و غیر پارامتری
- طراحی classifier (یادگیری با مربی)
  - رویکرد Bayesian یافتن بیشترین نزدیکی و شباهت
  - رویکرد جدا سازی
  - روش های مختلف تفسیر
- طراحی classifier (یادگیری بدون مربی)
  - خوشه بندی
- مسائل جنبی
  - تخمین میزان خطا و ریسک (خطای , Bayes و ...)
  - روش های ترکیب و بهبود کارآیی

مقدمه: تئوری احتمال و باز شناسی الگو

- دیدیم که در SPR اساس و پایه تفسیر و استدلال بر تئوری محاسبات احتمال قرار دارد.
- این تئوری بر این مبناست که گر چه در مواردی، نمی توان با دقت کامل از وقوع یا عدم وقوع پدیده ها سخن گفت و به عبارتی، به میزانی عدم قطعیت در کار است، اما این امر به معنای عدم وجود هیچ اطلاعاتی از پدیده ها نیست.
- تئوری احتمال با ارائه یک ساختار ریاضی، امکان تلفیق اطلاعات غیر قطعی را برای رساندن به برخی تصمیم گیری ها، فراهم می سازد.
- در این بحث، در مرحله اول نگاه حل مسئله با روش احتمال به مسائل باز شناسی الگو تبیین می شود.

بردار های اتفاقی

- در بحث بردار های اتفاقی، فرض بر این است که این نقاط (بردار ها) خروجی یک فرآیند تصادفی هستند که دارای توزیع خاصی است.
- معمولا توزیع های مختلف در رنج ورودی دارای مقادیر متفاوت بوده ولی این قابلیت وجود دارد که توسط تعداد معدودی پارامتر آنها را مشخص کرد.
- با مشخص بودن این توزیع، ایده های کلی مانند شباهت، وابستگی و فاصله برای خروجی فرآیند های فوق قابل تعریف است.
- این کمیت ها می توانند برای classification مورد استفاده قرار گیرند.

بررسی مجدد مثال تفکیک ماهی ها

- فرض کنیم در مسئله تفکیک ماهی ها، طول ماهی ها (که از روی تصویر بدست آمده است) را بعنوان مشخصه اصلی پذیرفته باشیم.
- با توجه به اینکه ممکن است ماهی ها ی یک نوع، در سنین مختلف و تحت شرایط رشد مختلف بوده باشند، قابل قبول است که هر یک از انواع دو گانه بتواند دارای طول های مختلفی باشد.
- به بیان دیگر طول یک ماهی، یک کمیت توام با عدم قطعیت (احتمالی) است و به طور قاطع نمی توان مقدار آن برای یک ماهی را داشت.
- از سوی دیگر، اگر چه طول هر دو نوع ماهی غیر قطعی است، اما غیر قطعی بودن آنها با هم یکسان نیست و همین امر خود می تواند مبنائی برای تفکیک دو نوع ماهی فراهم آورد.

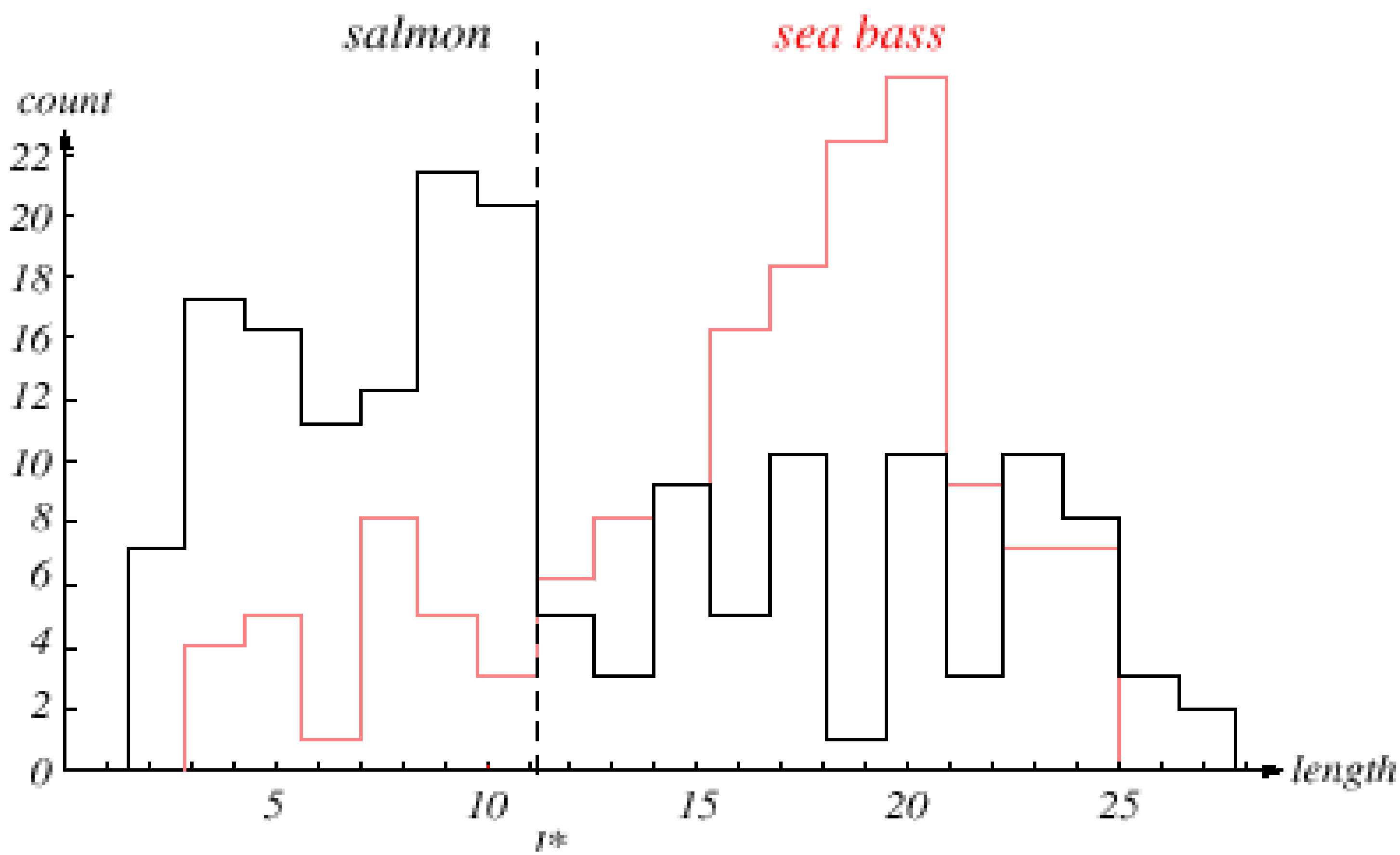
ادامه (یادگیری با مربی)

- ایده این است که ابتدا توسط سیستمی که می تواند جواب درست بدهد، (مثلا انسان ماهری که می تواند ماهی ها را از هم تفکیک کند) طبقه بندی روی تعدادی مثال صورت گیرد.
  - تشکیل مجموعه یادگیری (training set) که در آن هر نمونه شامل یک زوج مرتب است که درایه اول آن بردار مشخصه ها و درایه دوم آن برچسب یکی از کلاس هاست.(xj ,Wj)
- در هر یک ازکلاس ها یا طبقه ها(α)، مقادیر قابل سنجش، مشخصه ها وسپس چگونگی توزیع آن پیدا شود.

هیستوگرام، مبنائی برای توزیع احتمال

- می توان در نخستین گام، توزیع کمیت طول ماهی ها را برای هر یک از انواع آنها در طول های مختلف را، با هیستوگرام نشان داد
- برای این کار:
  - کمیت طول ماهی، به بازه هائی تقسیم می شود
  - تعداد ماهی هائی که (در هر نوع) مقدار طول آنها در هر یک از بازه ها قرار دارد پیدا می شود.
  - عدد مزبور(تعداد) به آن بازه نسبت داده می شود.
- اگر عدد مز بور را بر تعداد کل تقسیم کنیم (بطوری که سطح زیر منحنی ۱ شود)، می توان از این کمیت به احتمال اینکه طول ماهی مقدار بخصوصی باشد، تعبیر نمود.

هیستو گرام توزیع <u>طول</u> دو نوع ماهی



بررسی اولیه هیستو گرام

- اگر چه نمی توان برای هر یک از ماهی ها، از قبل طول دقیق را گفت، اما ملاحظه می شود که هر نوع، در طول های بخصوصی تعداد نمونه بیشتری دارد
- از این گذشته، این تعداد و محل وقوع حد اکثربرای دو نوع ماهی کاملا متفاوت است. مثلا:
  - سالمون بیشترین تعداد را در طول ۸ تا ۱۰
  - ماهی سی باس در طول ۱۸ تا ۲۲
- به بیان دیگر توزیع ماهی ها در طول های مختلف برای دو طبقه، متفاوت است.

بررسی و گرفتن ایده:

- بنظر می رسد که ماهی های salmon عموما طول کمتری دارند.(هر چند نه با قاطعیت بلکه با احتمال)
- ایده اساسی:
  - می توان از روش شمارش، به مفهوم احتمال رسید.
  - مثلا با شمارش تعدادماهی های یک نوع در یک بازه طولی مشخص (هیستو گرام) تخمینی از توزیع ماهی های این نوع به لحاظ طولی پیدا نمود.
- به نظر می آید که بتوان یک حد آستانه Thresholdیافت که برای طول های کمتر از آن، احتمال اینکه ماهی salmon باشد بیشتر و برای طول های کمتر از آن احتمال مزبور کمتر باشد.

طراحی کلاسیفایر، مرز تصمیم و ناحیه تصمیم

- این حد آستانه L یک <u>مرز تصمیم Decision Boundary</u> ارائه می کند.
- مرز تصمیم، فضای مشخصه ها را به دو ناحیه تصمیم Decision Region تقسیم می کند که هر ناحیه متناظر با یکی از کلاس هاست.
- اگر نقطه(نمونه ورودی) در ناحیه متعلق به کلاس اول بیفتد، ما آن نمونه را به آن کلاس منسوب خواهیم کرد و قس علیهذا
- بعبارت دیگر قاعده تفکیک یا classifier طرح شده می گوید:
  - اگر طول ماهی کوچکتر از L است ماهی salmon است و در غیر اینصورت sea bass می باشد.

کارآیی کلاسیفایر، احتمال خطا

- آیا با سیستم یاد شده، همه ماهی به نحو صحیح تفکیک می شوند؟
  خیر چرا که ممکن است باشند ماهی های salmon بزرگتر از L
  و sea bassکوچکتر از L
- آیا می توان گفت قانون ما در چند درصد موارد دچار خطا خواهد شد؟
- احتمال خطا برابر است با جمع دو احتمال شرطی یا احتمال مرکب:
  - احتمال اینکه ماهی salmon و ماهی بزرگتر از L باشد
  - احتمال اینکه ماهی sea bass باشد وماهی کوچکتر از L باشد

خلاصه:
استخراج قانون طبقه بندی بر اساس توزیع
ارزیابی قانون استخراج شده

- بر اساس یافته های فوق الذکر، این ایده به ذهن می رسد که سیستمی طرح شود تا بتواند طبقه بندی را بر اساس تفاوت دو نوع در توزیع انجام دهد.
  - مثلا تقسیم بازه طول به دو بخش که هر بخش به یکی از کلاس ها متعلق است
  - بخش اول بازه کوچکتر از ۱۰ و بزرگتر از ۲۵
  - بخش دوم بازه بزرگتر از ۱۰ و کوچکتر از ۲۵
- سپس بایستی کارآیی سیستم مزبور محاسبه شود
  - محاسبه تئوریک :اینکه سطح زیر منحنی توزیع هر کلاس در ناحیه کلاس دیگر چقدر است
  - محاسبه با تست سیستم روی نمونه ها : اینکه در تعداد معینی از نمونه های آزمایش، بخصوص نمونه هایی که اطلاعات آنها در مرحله یادگیری مورد استفاده قرار نگرفته اند، چقدر خطا رخ داده است یعنی برآورد قدرت تعمیم (generalization

استخراج قانون طبقه بندی
حال L را چگونه انتخاب کنیم؟

- نخستین چیزی که به ذهن می رسد این است که L باید به گونه ای انتخاب شود که ما دچار کمترین خطا شویم.
- به عبارتی می توانیم رابطه ای (فرمولی یا جدولی) بین مقدار احتمال خطا و آستانه L تشکیل دهیم (Error Surface) و نقطه (مرز)ای را بیابیم که خطا را می نیمم می کند.
- چنین رویکردی را <u>طراحی classifier برای مینیمم کردن خطای طبقه بندی</u> می گویند.
- ایده اصلی آن است که طول مورد نظر L را در مقادیر مختلف قرار داده و در هر ناحیه مقدار خطا را بسنجیم (بشماریم) و آن طولی را که کمترین خطا را موجی شده، به عنوان جواب (*) بپذیریم

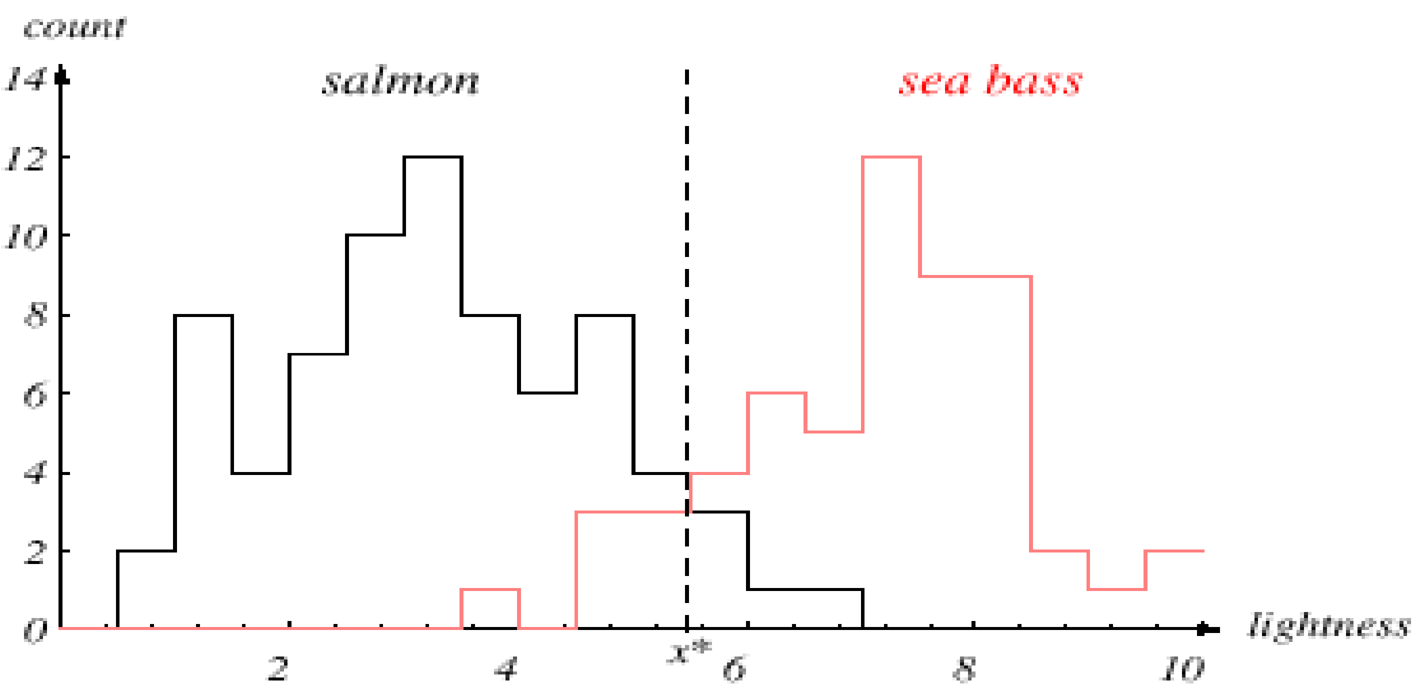
Error Vs. Risk

- سوال اساسی: آیا در مسائل عملی، همه خطا ها تاثیر یکسانی دارند؟
- فرض کنیم قیمت ماهی salmon بیش از sea bass باشد.
  - اگر بجای salmon، sea bass بگذاریم، پول بیشتری گرفته ایم
  - اگر تعدادمواردی که این خطا رخ می دهد زیاد شود، اعتبار ما کم شده و صدای مشتریان در می آید.
- بسته به اینکه کدامیک برای ما مهم تر است، ضرر(Risk) وارده به ما از دو نوع خطا ممکن است با هم مساوی نباشد. لذا می توان برای خطا ها وزن در نظر گرفت.
- در نظر گرفتن ضرر وارده، ما را به <u>طراحی بر اساس می نیمم کردن Risk</u> خواهد رسانید.

طراحی کلاسیفایر و بهینه سازی

- با توجه به مثال قبل، واضح است که فرآیند طراحی کلاسیفایر بر یک فرآیند بهینه سازی منطبق شده است.
- در چنین فرآیندی دو کار انجام گرفته است:
  - پیدا کردن معیاری که با اهداف ما سازگاری دارد (مانند معیار خطای طبقه بندی یا ریسک و هزینه طبقه بندی)
  - پیدا کردن روشی برای یافتن مقادیری از مشخصه ها، که معیار انتخاب شده را بهینه می کند
- اگر معیار مورد نظر، احتمال وقوع خطای انتساب باشد، و این احتمال از روی توابع توزیع احتمال به دست بیاید، کلاسیفایر مز بور به کلاسیفایر بیز Bayes موسوم است که مشخصه آن حداقل شدن احتمال خطای انتساب است

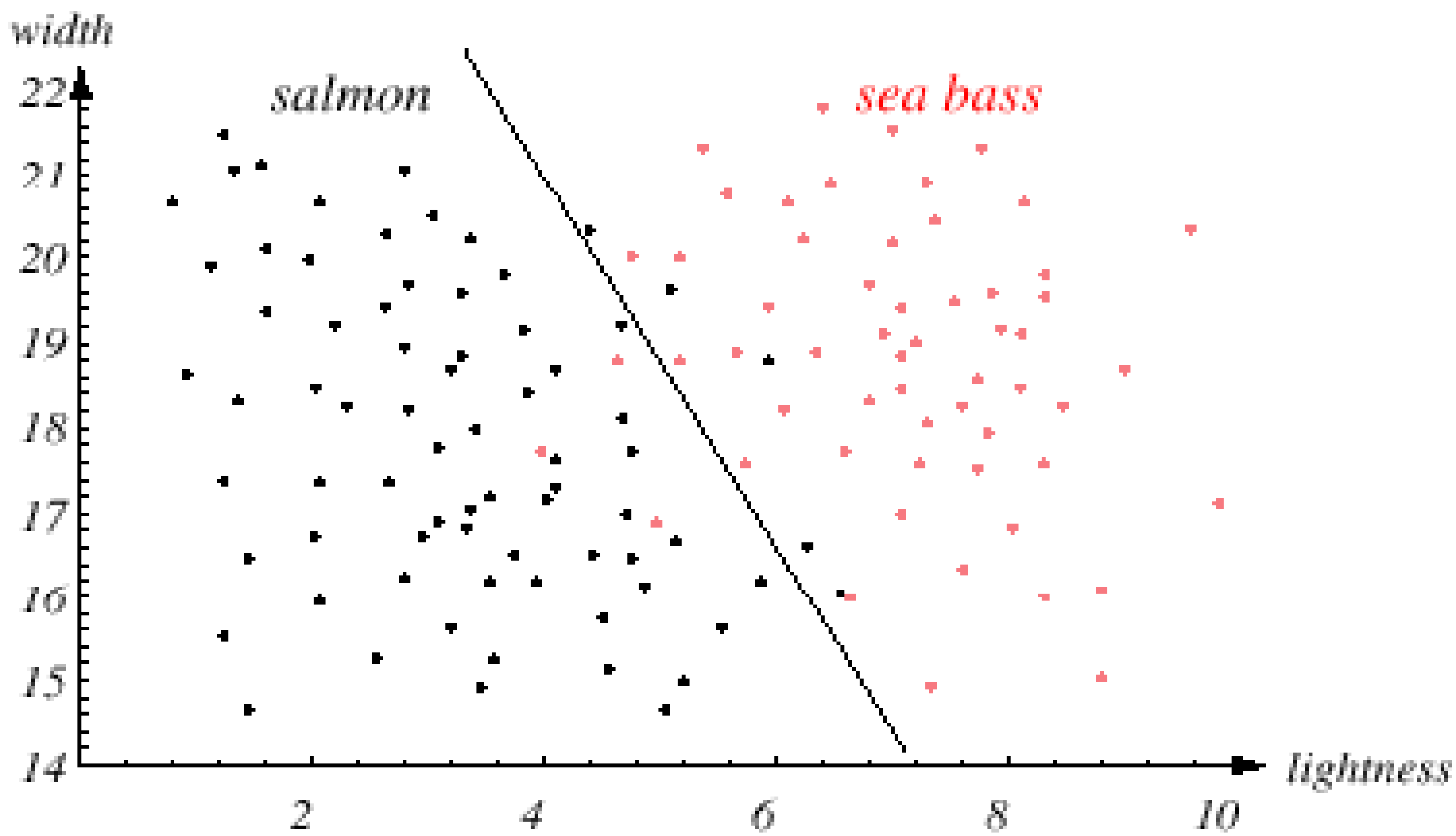
تاثیر مشخصه های انتخابی
یک مشخصه دیگر روشنی رنگ Lightness



بررسی مشخصه رنگ

- با یک بررسی بسیار اجمالی، مشخص می شود که مشخصه های مختلف، می توانند توزیع های مختلف و در نتیجه کارآئی های مختلفی را در طبقه بندی بوجود بیاورند
- به عنوان مثال مشخصه روشنی رنگ، احتمال خطای کمتری را بوجود می آورد اما ممکن است استفاده از آن، به دلیل نیاز به دوربین رنگی و پردازش های رنگی، هزینه بیشتری داشته و به همین دلیل در دستور کار قرار نگیرد.
- از سوی دیگر ممکن است اطلاعات بیشتری در دو مشخصه متفاوت وجود داشته باشد. لذا با استفاده از دو مشخصه، ممکن است بتوان به نتایج بهتری رسید.

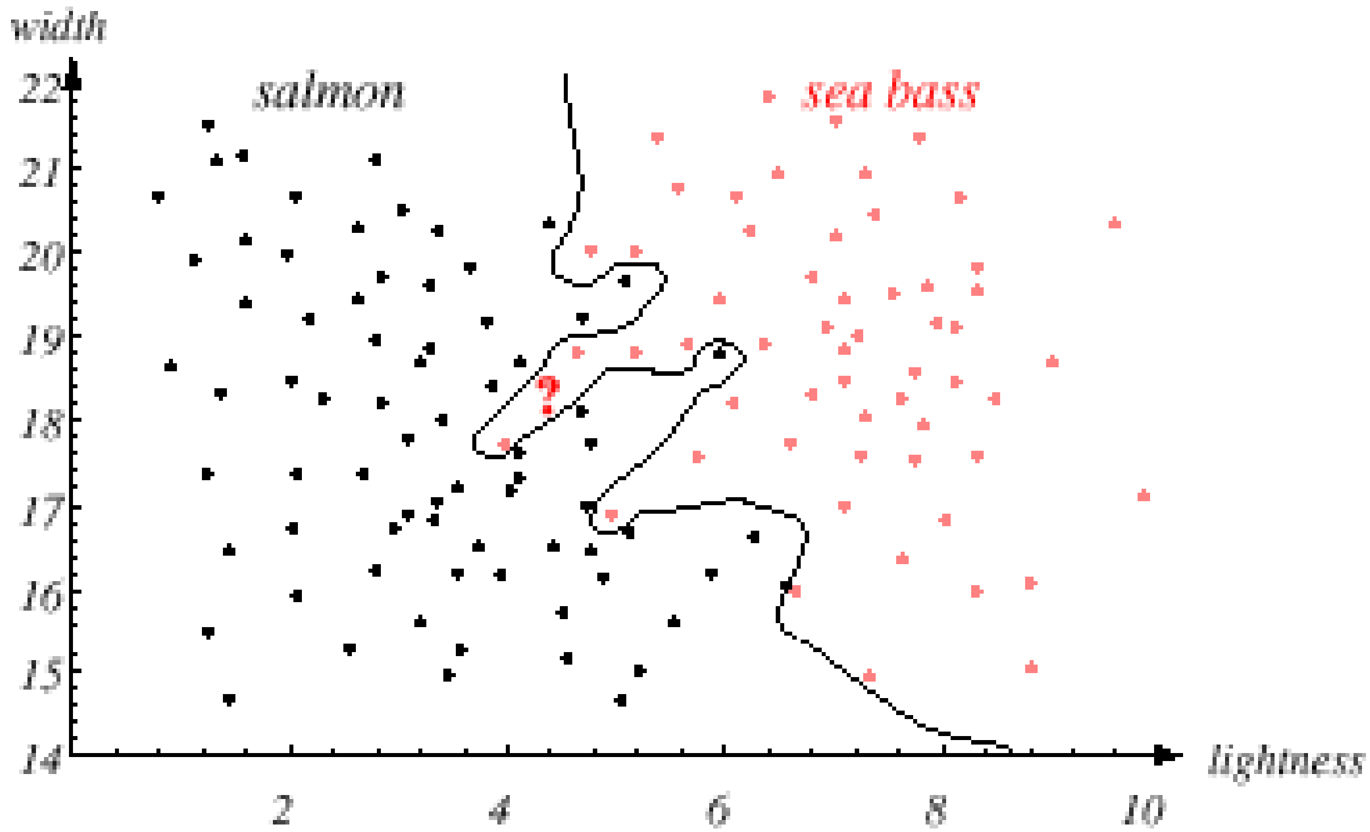
یک نمونه تصمیم دو بعدی بر اساس مشخصه روشنی:
مرز تصمیم دو بعدی(مثلا یک خط)



دستکاری اطلاعات، تفسیر

- اضافه شدن اطلاعات علی القاعده ممکن است قدرت تفکیکی بیشتر بوجود آورد.
- البته بدیهی است که هزینه هم بالاتر رفته است (احتمالا نوع سنجشگر ها و همچنین سرعت و قدرت پردازش مورد نیاز).
- آیا نمی شود از منحنی های دیگری بجز خط استفاده کرد تا به خطای کمتری رسید؟
- این امر در واقع به معنای آن است که در فرآیند طراحی کلاسیفایر، قیود دیگری گذاشته شود.

مرز تصمیم پیچیده تر با دو مشخصه و کلاسیفایر پیچیده تر و خطای کمتر



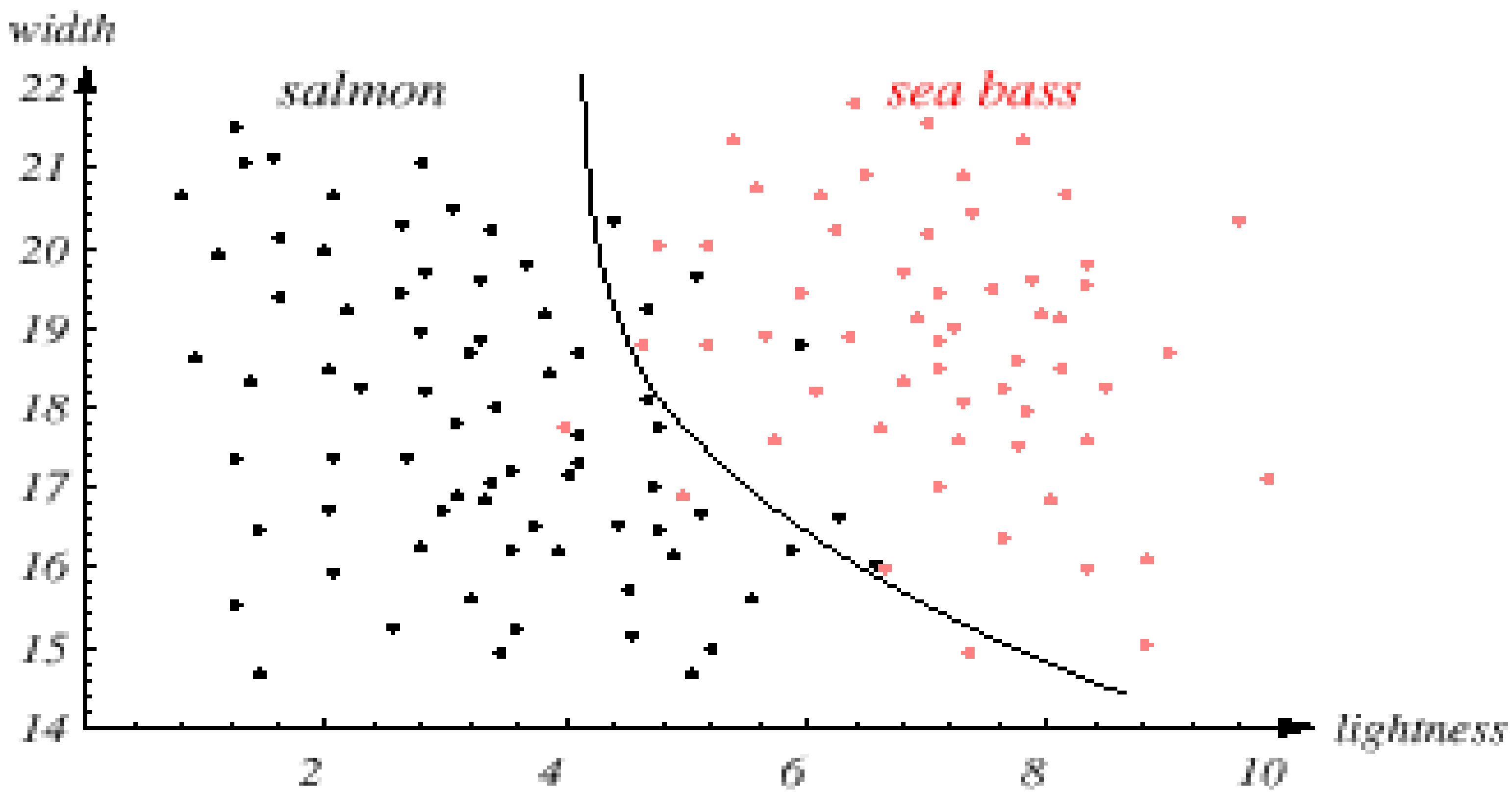
یادگیری در مقابل تعمیم

- نکته اساسی این است که بدلیل وجود خطا در سنجش ها، نقاط (بردار ها) مشخص کننده نمونه ها نه بردار های قطعی بلکه بردار های اتفاقی هستند.
- به همین دلیل معلوم نیست که کاهش دادن و به صفر رساندن خطا روی داده های یادگیری، باعث شود که سیستم روی بقیه داده ها هم خطای صفر را منجر شود.
- بعبارتی معلوم نیست یادگیری روی اطلاعات است یا روی نویز
- لذا بایستی سعی شود از پیچیدگی بیهوده سیستم جلوگیری گردد و قدرت تعمیم Generalization یعنی کارآیی روی داده هایی که سیستم بر اساس آنها ساخته نشده است، شکل گیرد.

training set & test set :ادامه

- برای رفع مشکل مزبور معمولا از همه نمونه ها برای طرح سیستم استفاده نمی شود بلکه بخشی از آن برای نشان دادن قابلیت تعمیم مورد استفاده قرار می گیرد
- به این بخش از داده که در واقع نتیجه کار سیستم طرح شده روی آن آزمایش می شود **test set** می گویند.
- از چند سیستم مختلف، سیستمی باید انتخاب شود که روی **test set** خطا(یا **risk**) کمتری داشته باشد حتی اگر روی مجموعه **training** خطای آن بیشتر باشد.

مرزتصمیم بهتر یا بهینه از نظر قابلیت تعمیم



خلاصه مسئله SPR

- Patternها یا نمونه ها با بردار مشخصه ها یا نقاط در صفحه مشخصه ها، بازنمایی می شوند.
- بدلیل عدم قطعیت موجود در اطلاعات، این نقاط در واقع بردار های اتفاقی هستند.
- برای طبقه بندی
  - به کمک اطلاعات بدست آمده از مجموعه یادگیری فضای مشخصه ها را می توان به نواحی متناظر با کلاس ها افراز نمود. این افراز به کمک مرز های تصمیم صورت می گیرد.
  - هر pattern ناشناخته جدید، می تواند با احتمالی، عضو هر یک از نواحی باشد.
  - Pattern مورد بحث را به کلاسی نسبت می دهیم که احتمال قرار داشتنش در ناحیه مربوط به آن کلاس، بیشترین باشد.

ادامه خلاصه SPR

- با این شیوه طبقه بندی، در اغلب موارد نتیجه تفسیر ما در واقع دارای درصدی از باور است و نه باور قطعی
- برای یافتن این درصد باور، آن را با این پیش فرض که معادل با احتمال است، از روی روش های آماری مورد محاسبه قرار می دهیم.
- از آنجائیکه معلوم نیست، داده هایی که در زمان یادگیری از آنها استفاده شده است، دقیقا همان خصوصیاتی را داشته باشند داده های دنیای واقعی دارند، لذاباید طرح روی داده هایی مورد آزمایش قرار گیرد که قبلا مورد استفاده قرار نگرفته اند.

ادامه خلاصه SPR

- لذا مراحل طراحی یک سیستم طبقه بندی اتوماتیک را می توان به شرح زیر در برشمرد:
  - اندازه گیری (انتخاب موارد قابل اندازه گیری)
  - پردازش های اولیه
  - بازنمایی یا استخراج و انتخاب مشخصه ها
  - طراحی سیستم کلاسیفایر (یادگیری)
  - آزمایش سیستم (تست)

بازشناسی آماری الگو

تعاریف و مفاهیم اولیه احتمالات

تجربه تعیینی و اتفاقی

- آزمایش(تجربه) اتفاقی: تجربه ای است که خروجی یا نتیجه آن بطور مطلق از قبل مشخص نیست و می تواند حالت های مختلفی داشته باشد.
  - انداختن سکه و تاس، مشغول بودن تلفن ...
- از آنجائیکه، نتیجه یک تجربه اتفاقی، نه یک مقدار یا حالت متعین بلکه شامل مجموعه ای از حالت هاست، بسیاری از مفاهیم، قوانین و فرمول های مر بوط به محاسبات اتفاقی با استفاده از ریاضیات و گزاره های مر بوط به مجموعه ها استخراج می شود.
- تعدا محدودی از از گزاره های مزبور غیر قابل اثبات و پذیرفتنی هستند و باقی موارد از روی آنها نتیجه می شوند.

فضای نمونه

- فضای نمونه Sample Space: مجموعه ای است که شامل همه حالت های مختلف از نتایج ممکن یک تجربه باشد.
  - مثال: برای تجربه انداختن تاس، فضای نمونه (گسسته)عبارتند از :

$$S = \{1, 2, 3, 4, 5, 6\}$$

- فضای نمونه ممکن است گسسته یا پیوسته باشد:
  - مثال : اگر زمان ارائه یک سرویس بین ۲ تا ۳ ساعت باشد، فضای نمونه (پیوسته) آن :

$$S = \{x \mid 2 \leq x \leq 3\}$$

  - مثال: اگر یک مشتری وارد یک سیستم صف شود که حداکثر ظرفیت آن ۵ نفر است، اگر او نداند که چند نفر پیش از او در صف هستند، فضای نمونه برای طول صف پیش از او عبارت خواهد بوداز : S={1,2,3,4}

تعاریف و مفاهیم...

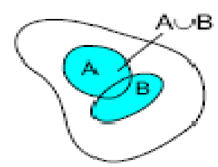
- **<u>پیشامد یا رویداد (Event)</u>** : به هر زیر مجموعه از فضای نمونه، یک پیشامد می گویند.
- اکثر گزاره ها و احکامی که در احتمالات مورد توجه قرار می گیرد در مورد پیشامد هاست و می توان آنها را به سه دسته تقسیم کرد:
  - <u>پیشامد های قطعی</u>، شامل همه فضای نمونه هستند،
  - <u>پیشامد های غیر ممکن</u> شامل مجموعه تهی است،
  - <u>پیشامد های محتمل</u>: بین دو حالت فوق هستند.
- مثال : پیشامد فرد بودن نتیجه انداختن تاس: $A=\{1,3,5\}$

ترکیب پیشامد ها:
(از تئوری مجموعه ها)

Combining events

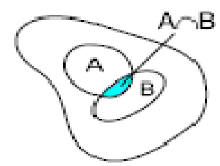
Union "A or B".

$$A \cup B = \{e \in \mathcal{S} \mid e \in A \text{ or } e \in B\}$$



Intersection (joint event) "A and B".

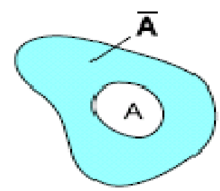
$$A \cap B = \{e \in \mathcal{S} \mid e \in A \text{ and } e \in B\}$$



Events A and B are mutually exclusive, if $A \cap B = \phi$.

Complement "not A".

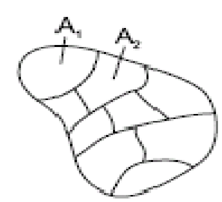
$$\bar{A} = \{e \in \mathcal{S} \mid e \notin A\}$$



Partition of the sample space

A set of events $A_1, A_2, \ldots$ is a partition of the sample space $\mathcal{S}$ if

1. The events are mutually exclusive, $A_i \cap A_j = \phi$, when $i \neq j$.
2. Together they cover the whole sample space, $\cup_i A_i = \mathcal{S}$.



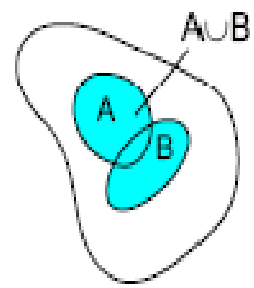
احتمال وقوع پیشامد، تابع احتمال

- برای هر پیشامد A، بر اساس اصل شمارش به طور تجربی، احتمال وقوع ، به شرح زیرتعریف می شود:
  - احتمال A عبارت است از نسبت تعداد وقوع حالت های پیشامد A به تعدادکلحالت ها
  - از آنجائیکه احتمال مزبور متغیر بوده و به A بستگی دارد آن را تابع تعریف می کنند:

$$P\{A\} = \lim_{N \to \infty} N(A)/N \qquad \begin{cases} N & = \text{number of experiments} \\ N(A) & = \text{number of occurrences of } A \end{cases}$$

- در برخی موارد با اینکه ممکن است شمارش واقعی (بخصوص برای بی نهایت مرتبه) صورت نگرفته باشد، بر مبنای استدلال می توان احتمال را بدست آورد.
  - مانند زمانی که می گوئیم احتمال اینکه نتیجه انداختن یک تاس ۱ باشد، یک ششم است.

خواص تابع احتمال



1. $0 \leq \mathrm{P}\{A\} \leq 1$

2. $\mathrm{P}\{\mathcal{S}\} = 1 \quad \mathrm{P}\{\phi\} = 0$

3. $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\} - \mathrm{P}\{A \cap B\}$

4. If $A \cap B = 0$, then $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\}$

   If $A_i \cap A_j = 0$ for $i \neq j$, then $\mathrm{P}\{\cup_i A_i\} = \mathrm{P}\{A_1 \cup \ldots \cup A_n\} = \mathrm{P}\{A_1\} + \ldots \mathrm{P}\{A_n\}$

5. $\mathrm{P}\{\bar{A}\} = 1 - \mathrm{P}\{A\}$

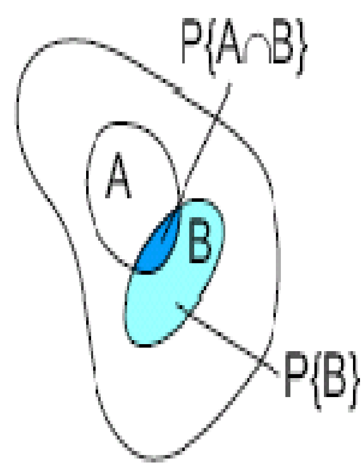
6. If $A \subseteq B$, then $\mathrm{P}\{A\} \leq \mathrm{P}\{B\}$

تابع احتمال شرطی (مشترک):
Joint or conditional probability

- این تابع نشان دهنده احتمال این است که با فرض وقوع B، پیشامد A اتفاق بیفتد و طبق اصل شمارش مقدار آن برابر است با نسبت حالت های مشترک A با B با کل حالت های B:

The probability of event A given that B has occurred.

$$\boxed{P\{A|B\} = \frac{P\{A \cap B\}}{P\{B\}}} \qquad \Rightarrow P\{A \cap B\} = P\{A|B\}P\{B\}$$



قانون جمع احتمال ها:

- اگر مجموعه پیشامد های B_i، کل فضای نمونه را افرازکند می توان احتمال A را با استفاده از احتمال مشترک آن با B_i هایافت. یعنی:

$$\begin{array}{llllll} 1. & \cup_i B_i & = & \mathcal{S} & \text{certain event} & \mathrm{P}\{\cup_i B_i\} = 1 \\ 2. & B_i \cap B_j & = & \phi & \text{for } i \neq j & \mathrm{P}\{B_i \cap B_j\} = 0 \end{array}$$

Then $A = A \cap \mathcal{S} = A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$ and

$$\boxed{\mathrm{P}\{A\} = \sum_{i=1}^{n} \mathrm{P}\{A \cap B_i\} = \sum_{i=1}^{n} \mathrm{P}\{A|B_i\}\mathrm{P}\{B_i\}}$$

فرمول Bayes:

- این فرمول برای محاسبه احتمال پسین مشترک یکی از حالت ها (پس از مشاهده) تدوین گردیده است و می گوید:
- اگر مجموعه پیشامد های B_i فضای نمونه را افراز کرده باشد، وبا فرض وقوع پیشامد A بخواهیم احتمال شرطی هر یک از پیشامد های آن مجموعه را محاسبه کنیم، خواهیم داشت:

$$\boxed{\mathrm{P}\{B_i|A\} = \frac{\mathrm{P}\{A \cap B_i\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{A|B_i\}\mathrm{P}\{B_i\}}{\sum_j \mathrm{P}\{A|B_j\}\mathrm{P}\{B_j\}}}$$

فرمول Bayes:

- نامگذاری کمیت های فرمول Bayes:

$P\{B_i \mid A\}$ posterior probability

$p\{A|B_i\}$ likelihood or distribution

$P\{B_i\}$ prior probability

$\sum_j P\{A \mid B_i\}.P\{B_i\}$ normalising term

فرمول Bayes ...مثال

- در یک جامعه ۴۵٪جمعیت مردان و ۵۵٪ زنان هستند.
- ۴۰٪ زنان و ۷۵٪مردان شاغلند.
- می خواهیم احتمال اینکه یک فرد شاغل، زن باشد را بیابیم:

پیشامد مرد یا زن بودن را با B_i و شاغل بودن را با A نشان می دهیم:

$P(B_1) = 0.55 \quad P(B_2) = 0.45$

$P(A|B_1) = 0.4 \quad P(A|B_2) = 0.75$

$$P(B_1|A) = \frac{P(A|B_1).P(B_1)}{P(A|B_1).P(B_1) + P(A|B_2).P(B_2)}$$

$$= \frac{0.4 \times 0.55}{0.4 \times 0.55 + 0.75 \times 0.45} = 0.504$$

استقلال پیشامد ها:

- دو پیشامد را مستقل از هم می گوئیم هر گاه وقوع یکی از آنها، تاثیری در احتمال وقوع دیگری نداشته باشد.
- به تعبیر دیگر: احتمال وقوع شرطی (مشترک) دو پیشامد مستقل، برابر با حاصلضرب احتمال وقوع هر یک از آنها باشد:
- بین مستقل بودن وانحصاری بودن تفاوت است

$$P\{A \cap B\} = P\{A\}.P\{B\}$$

$$P\{A|B\} = \frac{P\{A \cap B\}}{P\{B\}} = P\{A\}$$

پیشامد های مستقل... مثال

Example 1: Tossing two dice, $A = \{n_1 = 6\}$, $B = \{n_2 = 1\}$

$$A \cap B = \{(6,1)\}, \quad P\{A \cap B\} = \tfrac{1}{36}, \qquad \text{all combinations equally probable}$$

$$P\{A\} = P\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} = \tfrac{6}{36} = \tfrac{1}{6}; \quad \text{similarly } P\{B\} = \tfrac{1}{6}$$

$$P\{A\}P\{B\} = \tfrac{1}{36} = P\{A \cap B\} \quad \Rightarrow \quad \text{independent}$$

Example 2: $A = \{n_1 = 6\}$, $B = \{n_1 + n_2 = 9\} = \{(3,6),(4,5),(5,4),(6,3)\}$

$$A \cap B = \{(6,2)\}$$

$$P\{A\} = \tfrac{1}{6}, \quad P\{B\} = \tfrac{4}{36}, \quad P\{A \cap B\} = \tfrac{1}{36}$$

$$P\{A\} \cdot P\{B\} \neq P\{A \cap B\} \quad \Rightarrow \quad A \text{ and } B \text{ dependent}$$

متغیر های تصادفی و توزیع:
Random Variables and Distribution

- در مسائل عملی با تجربه های اتفاقی، معمولا بیش از آن که به چگونگی وقوع خروجی ها علاقمند باشیم، به تعداد دفعات وقوع آنها توجه داریم.
  - مثلا در یک تجربه تکرار انداختن سکه، برایمان ترتیب شیر یا خط مهم نیست بلکه می خواهیم بدانیم چند بار شیر و چند بار خط می آید.
- متغیر اتفاقی حقیقی X، نگاشتی از فضای نمونه S به فضای حقیقی R فراهم می کند که به هر پیشامد e Є S، عددی مانند X(e) نسبت می دهد(که مثلا نشانگر احتمال وقوع آن پیشامد است).
- تصویر (image) یک متغیر اتفاقی عبارتست از: مجموعه همه اعدادی که آن متغیر می تواند بگیرد(برد نگاشت)

$$S_X = \{x \in R \mid X(e) = x, e \in S\}$$

مثال

- اگر در یک تجربه ۳ بار پرتاب سکه، عدد $X(e)$ ، نشان دهنده تعداد آمدن شیر(یا خط) باشد، تصویر این متغیر اتفاقی، مجموعه {0,1,2,3} است (متغیر گسسته اتفاقی)

e	$X(e)$
hhh	3
hht	2
hth	2
htt	1
thh	2
tht	1
tth	1
ttt	0

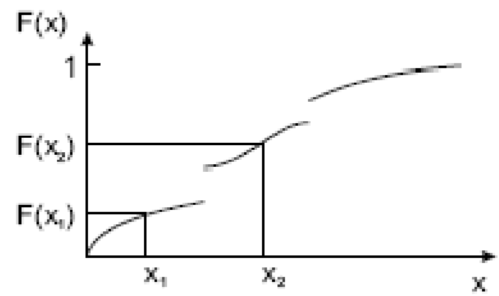
- The values of X are "drawn" by "drawing" e
- e represents a "lottery ticket", on which the value of X is written

تابع توزیع جمعی، تابع چگالی توزیع و تابع دنباله(برای متغیر اتفاقی پیوسته)
(cdf=cumulative distribution function) & (pdf=probability density function)

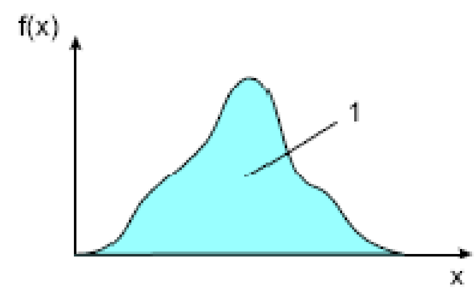
$$\boxed{F(x) = \mathrm{P}\{X \leq x\}}$$

The probability of an interval

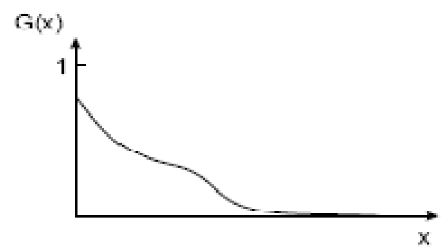
$$\boxed{\mathrm{P}\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)}$$



$$\boxed{f(x) = \frac{dF(x)}{dx} = \lim_{dx \to 0} \frac{\mathrm{P}\{x < X \leq x + dx\}}{dx}}$$



Complementary distribution function (tail distribution)

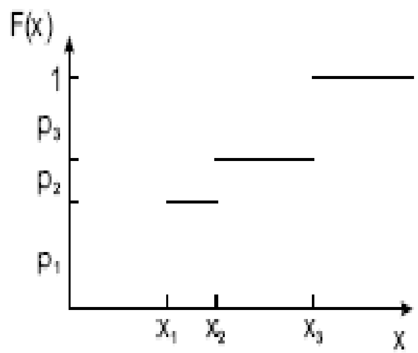


$$\boxed{G(x) = 1 - F(x) = \mathrm{P}\{X > x\}}$$

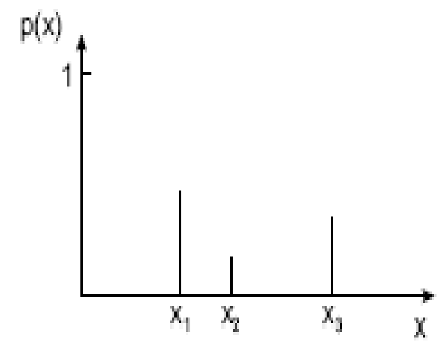
تابع جرم (برای متغیر اتفاقی گسسته)

- این متغیر ها تنها برخی از مقادیر(مجزا) را می توانند اختیار نمایند لذا پیوستگی در مقادیر آنها وجود ندارد.
- و برای این گونه متغیر ها، تابع جرم تعریف می کنند:

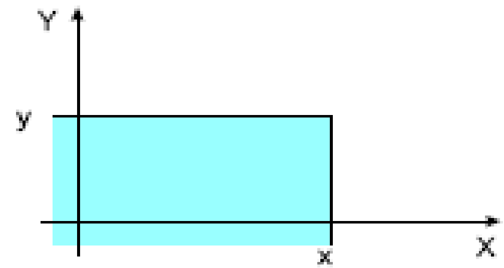
(pmf=probability mass function)



$$p(x) = \mathrm{P}\{X = x\} = \begin{cases} p_i & \text{when } x = x_i \\ 0, & \text{otherwise} \end{cases}$$



تابع توزیع و جرم دوتایی (مشترک) یا شرطی، شرط استقلال آنها

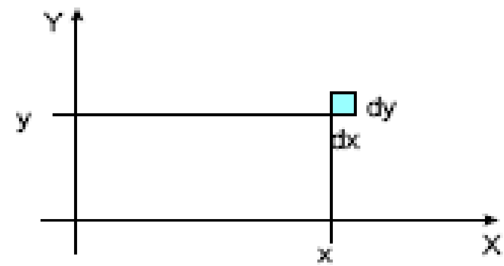


Joint distribution function

$$\boxed{F_{X,Y}(x,y) = \mathrm{P}\{X \le x, Y \le y\}}$$

Joint probability density function

$$\boxed{f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)}$$



the events $\{X \le x\}$ and $\{Y \le y\}$ are independent, whence

$$\boxed{F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)}$$

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$$

عملگر هاو پارامتر های مشخص کننده تابع توزیع:

- معمولا به دو دلیل برای توابع توزیع پارامتر های مشخص کننده تعریف می کنیم:
  - نخست اینکه توابع عملی ممکن است روابط ریاضی پیچیده ای داشته باشند و گاها برای درک بهتر لازم باشد آنها را به شکل فشرده تر نمایش دهیم.
  - دوم اینکه از آنجائیکه مقدار عددی متغیر های اتفاقی کاملا معلوم نیست، حتی اگر تابع توزیع آنها را داشته باشیم، تعیین مقدار دقیق در هر نقطه (آرگومان) ممکن نیست. اما برای تعداد نسبتا زیاد تجربه ها، می توان برخی مقادیر را که نشان دهنده رفتار است تعیین نمود.
- پارامتر های مشخص کننده با استفاده از عملگر های خاص (روی همه توابع توزیع) تعریف می شوند.

expectation امید ریاضی

Denoted by $\mathrm{E}[X] = \bar{X}$

Continuous distribution: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx}$

Discrete distribution: $\boxed{\mathrm{E}[X] = \sum_i x_i p_i}$

In general: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, dF(x)}$ $dF(x)$ is the probability of the interval dx

Properties of expectation

$\mathrm{E}[cX] = c\mathrm{E}[X]$	c constant
$\mathrm{E}[X_1 + \cdots X_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]$	always
$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$	only when X and Y are independent

واریانس

Denoted by V[X] (also Var[X])

$\boxed{\mathrm{V}[X] = \mathrm{E}[(X - X)^2]} = \mathrm{E}[X^2] - \mathrm{E}[X]^2$

Covariance

Denoted by Cov[X, Y]

$\boxed{\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \bar{X})(Y - \bar{Y})]} = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$

$\mathrm{Cov}[X, X] = \mathrm{V}[X]$

If X are Y independent then $\mathrm{Cov}[X, Y] = 0$

Properties of variance

$\mathrm{V}[cX] = c^2\mathrm{V}[X]$	c constant; observe square
$\mathrm{V}[X_1 + \cdots X_n] = \sum_{i,j=1}^{n} \mathrm{Cov}[X_i, X_j]$	always
$\mathrm{V}[X_1 + \cdots X_n] = \mathrm{V}[X_1] + \cdots + \mathrm{V}[X_n]$	only when the X_i are independent

Properties of covariance

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$$

$$\mathrm{Cov}[X + Y, Z] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

امید ریاضی شرطی

$$\boxed{\mathrm{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y)dx}$$

Properties of conditional expectation

$\mathrm{E}[X \mid Y] = \mathrm{E}[X]$	if X and Y are independent
$\mathrm{E}[cX \mid Y] = c\,\mathrm{E}[X \mid Y]$	c is constant
$\mathrm{E}[X + Y \mid Z] = \mathrm{E}[X \mid Z] + \mathrm{E}[Y \mid Z]$	
$\mathrm{E}[g(Y) \mid Y] = g(Y)$	
$\mathrm{E}[g(Y)X \mid Y] = g(Y)\mathrm{E}[X \mid Y]$	

توزیع نرمال یا گوسی

- توزیع گوسی از چند جهت در بحث بازشناسی آماری مهم است:
  - بسیاری از پدیده های طبیعی دارای مشخصاتی شبیه این توزیع هستند.
  - خصوصیات ریاضی این توزیع باعث به دست آمدن روابط ساده ای می شود که خصوصیات این تکنیک را به خوبی مشخص می کند.

$$p(x|\omega_i)=\frac{1}{(2\pi)^{d/2}.\Sigma^{1/2}}\exp[\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$$

x	input vector
μ	mean vector
Σ	covariance matrix
d	dimention of the x
$p(x\mid\omega_i)$	prbability distribution function

خلاصه بحث:

- در این جلسه دو بحث اصلی ارائه شد:
  - چگونگی دیدگاه حل مسئله در SPR و مراحل اصلی
    - چگونگی امکان اخذ تصمیم بر اساس توابع احتمال و توزیع آن
  - یادآوری مفاهیم اولیه آمار و احتمالات
    - معرفی متغیر تصادفی و توابع توزیع احتمال
    - فرمول Bayes در ترکیب اطلاعات توابع توزیع
    - معرفی پارامتر های اصلی توابع توزیع احتمال