

جواب سوال 15 و 16)

(۱۵) تفاوت بین روش های پارامتریک و غیر پارامتری به شرح زیر است :

ردیف	روش های پارامتریک	روشهای غیر پارامتریک
۱	روش های پارامتری از تعداد مشخصی پارامتر برای ساخت مدل استفاده می کنند.	روش های غیر پارامتری از تعداد انعطاف پذیر پارامترها برای ساخت مدل استفاده می کنند.
۲	تجزیه و تحلیل پارامتری برای آزمایش میانگین گروه است.	تجزیه و تحلیل غیر پارامتری برای آزمایش میانگین ها است.
۳	فقط برای متغیرها قابل اجراست.	برای هر دو قابل استفاده است – متغیر و ویژگی.
۴	همیشه فرضیات قوی در مورد داده ها را در نظر می گیرد.	به طور کلی فرضیات کمتری در مورد داده ها وجود دارد.
۵	روش های پارامتریک نسبت به روش های غیر پارامتریک به داده های کمتری نیاز دارد.	روش های غیر پارامتریک به داده های بسیار بیشتری نسبت به روش های پارامتریک نیاز دارد.
۶	روش های پارامتری فرض می شود که یک توزیع طبیعی باشد.	توزیع فرضی در روش های غیر پارامتری وجود ندارد.
۷	دستگیره های داده های پارامتری – داده های فاصله ای یا داده های نسبت	اما روش های غیر پارامتری داده های اصلی را کنترل می کنند.
۸	در اینجا وقتی از روشهای پارامتریک استفاده می کنیم، نتیجه یا خروجی های تولید شده به راحتی می توانند تحت تأثیر عوامل دور از دسترس قرار بگیرند.	هنگامی که ما از روش های غیر پارامتریک استفاده می کنیم، نتیجه یا خروجی های تولید شده نمی توانند به طور جدی تحت تأثیر عوامل دور از انتظار قرار بگیرند.
۹	روش های پارامتریک می توانند در بسیاری از شرایط عملکرد خوبی داشته باشند اما هنگامی که گسترش هر گروه متفاوت باشد عملکرد آن در اوج (بالا) است.	به همین ترتیب، روش های غیر پارامتریک می توانند در بسیاری از شرایط عملکرد خوبی داشته باشند اما وقتی گسترش هر گروه یکسان باشد، عملکرد آن در اوج (بالا) است.
۱۰	قدرت آماری روشهای پارامتریک بیشتر از روشهای غیر پارامتریک است.	روشهای غیر پارامتریک از نظر آماری قدرت کمتری نسبت به روشهای پارامتریک دارند.
۱۱	تا آنجا که محاسبه در نظر گرفته می شود، این روش ها از نظر محاسباتی سریعتر از روش های غیر پارامتری هستند.	تا آنجا که محاسبه در نظر گرفته می شود، این روش ها از نظر محاسباتی کندتر از روش های پارامتری هستند.
۱۲	مثال ها- رگرسیون لجستیک، مدل Naïve Bayes و غیره	مثال ها- KNN، مدل درخت تصمیم گیری و غیره

۱۶) معیار Matthews Correlation Coefficient (MCC)

کدام معیارها باید بهتر باشند؟

بدیهی است معیار های که تشخیص درست را نمایش میدهند باید بهتر باشند که در زیر با رنگ سبز نشان داده شده است.

۱- مثبت صحیح (True Positive) = درست شناسایی شده است.

۲- مثبت کاذب (False Positive) = اشتباه شناسایی شده است (خطای نوع یک در انجام آزمون).

۳- منفی صحیح (True Negative) = به درستی رد شد.

۴- منفی کاذب (False Negative) = اشتباه رد شد (خطای نوع دوم در انجام آزمون).

معیارهای ارزیابی MCC:

پارامتر دیگری است که برای ارزیابی کارایی الگوریتم های یادگیری ماشین از آن استفاده می شود. این پارامتر بیان گر کیفیت کلاس بندی برای یک مجموعه باینری می باشد. MCC (Matthews correlation coefficient)، سنجی ای است که بیان گر بستگی مابین مقادیر مشاهده شده از کلاس باینری و مقادیر پیش بینی شده از آن می باشد. مقادیر مورد انتظار برای این کمیت در بازه ۱- و ۱+ متغیر می باشد. مقدار ۱+، نشان دهنده پیش بینی دقیق و بدون خطای الگوریتم یادگیر از کلاس باینری می باشد. مقدار ۰، نشان دهنده پیش بینی تصادفی الگوریتم یادگیر از کلاس باینری می باشد. مقدار ۱-، نشان دهنده عدم تطابق کامل مابین موارد پیش بینی شده از کلاس باینری و موارد مشاهده شده از آن می باشد. مقدار این پارامتر را به طور صریح، با توجه به مقادیر ماتریس آشفستگی به شرح زیر، می توان محاسبه نمود:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

معیارهای ارزیابی MCC