



MINDCRAFT

CRAFTING INTELLIGENT MINDS

Disclaimer: This content is generated by AI.

मशीन लर्नगि में बुनियादी एल्गोरदिम

:

इस मॉड्यूल में, शकिषार्थियों को रैखिक प्रतगिमन, नरिणय वृक्ष और -नकिटतम पड़ोसियों जैसे मौलिक मशीन लर्नगि एल्गोरदिम से परिचित कराया जाएगा। मॉड्यूल इन एल्गोरदिम और सरल उपयोग मामलों के पीछे के सिद्धांतों पर ध्यान केंद्रित करता है।

मशीन लर्नगि एल्गोरदिम का परिचय

मशीन लर्नगि आर्टिफिशियल इंटेलिजेंस का एक अभिन्न क्षेत्र है जिसमें एल्गोरदिम और सांख्यिकीय मॉडल शामिल हैं जो कंप्यूटर को स्पष्ट निर्देशों के बिना कार्य करने में सक्षम बनाते हैं। यह उप-मॉड्यूल मौलिक मशीन लर्नगि एल्गोरदिम का एक व्यापक अवलोकन प्रदान करता है, जिसमें मुख्य रूप से पर्यवेक्षण, अप्रशिक्षित और सुदृढीकरण सीखने में वर्गीकृत किया गया है, साथ ही उनके अनुप्रयोगों और वास्तविक दुनिया की प्रासंगिकता के साथ।

1. मशीन लर्नगि एल्गोरदिम की परिभाषा

मशीन लर्नगि एल्गोरदिम एक ऐसी प्रक्रिया है जो कंप्यूटर को डेटा से सीखने और उसके आधार पर भविष्यवाणियां या निर्णय लेने की अनुमति देती है। एल्गोरदिम पैटर्न की पहचान करते हैं, वर्गीकरण, प्रतगिमन, क्लस्टरिंग और बहुत कुछ सक्षम करते हैं। उन्हें सीखने के कार्य की प्रकृति के आधार पर विभिन्न प्रकारों में वर्गीकृत किया जा सकता है, जैसे कि पर्यवेक्षण, अप्रशिक्षित और सुदृढीकरण सीखना।

2. मशीन लर्नगि एल्गोरदिम के प्रकार

एल्गोरदिम को मुख्य रूप से तीन श्रेणियों में वर्गीकृत किया जा सकता है: - **पर्यवेक्षण शिक्षण** : ये एल्गोरदिम लेबल किए गए डेटा से सीखते हैं। इनपुट-आउटपुट जोड़े का उपयोग करके एक मॉडल को

प्रशिक्षित किया जाता है, जिससे यह अनदेखे डेटा के लिए आउटपुट की भविष्यवाणी कर सकता है।
उदाहरण: एक सामान्य एल्गोरिदम रैखिक प्रतगमन एल्गोरिदम है जिसका उपयोग आकार, स्थान आदि जैसी विभिन्न विशेषताओं के आधार पर घर की कीमतों की भविष्यवाणी करने के लिए किया जाता है। यदि किसी डेटासेट में कई घरों के लिए वर्ग फुटेज और मूल्य टैग शामिल हैं, तो एक पर्यवेक्षित शिक्षण मॉडल नए घरों की कीमतों की भविष्यवाणी करने के लिए सबसे उपयुक्त रेखा स्थापित कर सकता है। -
****अनपर्यवेक्षित शिक्षण****: यहां, परिणामों के पूर्व ज्ञान के बिना पैटर्न या समूहों की खोज करने के लिए एल्गोरिदम को लेबल रहित डेटा पर लागू किया जाता है। ***उदाहरण***: -मैनस क्लस्टरिंग एक अनपर्यवेक्षित एल्गोरिदम है जो समान डेटा बिंदुओं को एक साथ समूहित करता है। उदाहरण के लिए, ग्राहक विभाजन में, व्यवसाय पूर्व-निर्धारित लेबल के बिना क्रय व्यवहार के आधार पर अलग-अलग ग्राहक समूहों की पहचान करने के लिए -मैनस का उपयोग कर सकते हैं। - ****सुदृढीकरण सीखना****: इसमें ऐसे एल्गोरिदम शामिल हैं जो गतिशील वातावरण में अपने कार्यों से प्रतिक्रिया प्राप्त करके सीखते हैं, समय के साथ परीक्षण और त्रुटि के माध्यम से अपने प्रदर्शन को अनुकूलित करते हैं। ***उदाहरण***: एक सुदृढीकरण सीखने के एल्गोरिदम को गेम खेलने में लागू किया जा सकता है (जैसे अलफागो) जहां मॉडल खुद के खिलाफ प्रतस्पर्धा करके और जीत या हार के आधार पर अपनी रणनीति को परिष्कृत करके इष्टतम रणनीति सीखता है।

3. मशीन लर्निंग एल्गोरिदम में प्रमुख अवधारणाएँ

मशीन लर्निंग एल्गोरिदम किस तरह से प्रभावी ढंग से काम करते हैं, इसे समझने के लिए कुछ प्रमुख अवधारणाओं को समझना ज़रूरी है: - ****प्रशिक्षण और परीक्षण डेटासेट****: अपने डेटासेट को दो सेट में विभाजित करें: एक प्रशिक्षण सेट जिससे एल्गोरिदम सीखता है, और एक परीक्षण सेट जो सत्यापित करता है कि मॉडल अज्ञात डेटा पर कितना अच्छा प्रदर्शन करता है। - ****ओवरफिटिंग और अंडरफिटिंग****: ओवरफिटिंग तब होती है जब कोई मॉडल प्रशिक्षण डेटा को बहुत अच्छी तरह से सीखता है, जिसमें शोर और आउटलेयर शामिल हैं, जिससे यह नए डेटा पर खराब प्रदर्शन करता है। अंडरफिटिंग तब होती है जब कोई मॉडल डेटा की अंतर्निहित प्रवृत्तियों को पकड़ने के लिए बहुत सरल होता है। - ****मूल्यांकन मेट्रिक्स****: सटीकता, प्रशुद्धता, रिकॉल और 1 स्कोर जैसे विभिन्न मेट्रिक्स का उपयोग मॉडल के प्रदर्शन का मूल्यांकन करने के लिए किया जाता है, जो विभिन्न एल्गोरिदम की प्रभावशीलता को निर्धारित करने में सहायता करता है।

4. मशीन लर्निंग एल्गोरिदम के वास्तविक-विश्व अनुप्रयोग

मशीन लर्निंग एल्गोरिदम विभिन्न उद्योगों में महत्वपूर्ण भूमिका निभाते हैं: - ****स्वास्थ्य सेवा****: एल्गोरिदम रोगी के इतिहास और पैटर्न के आधार पर बीमारियों की पहचान करने के लिए पूर्वानुमानात्मक विश्लेषण में मदद करते हैं। - ****वित्त****: क्रेडिट स्कोरिंग एल्गोरिदम आवेदकों के वित्तीय इतिहास के आधार पर उन्हें ऋण देने के जोखिम का आकलन करते हैं। - ****खुदरा****: अनुशंसा प्रणालियाँ उपयोगकर्ताओं को उनकी पछिली खरीदारी और व्यवहार पैटर्न के आधार पर उत्पादों का सुझाव देने के लिए सहयोगी फिल्टरिंग एल्गोरिदम का उपयोग करती हैं।

5. मशीन लर्निंग एल्गोरिदम का निष्कर्ष और भविष्य

मशीन लर्निंग एल्गोरिदम लगातार विकसित हो रहे हैं, जिससे अनगिनत क्षेत्रों में नवाचार को बढ़ावा मिल रहा है। जैसे-जैसे अधिक डेटा उपलब्ध होगा और कंप्यूटेशनल शक्ति बढ़ेगी, इन एल्गोरिदम की क्षमताएं और अनुप्रयोग बढ़ेंगे, जिससे उद्योगों के संचालन और सेवाएं प्रदान करने के तरीके में मौलिक रूप से बदलाव आएगा।

Reference:

<https://www.ibm.com/cloud/learn/machine-learning>

<https://scikit-learn.org/stable/about.html>

<https://towardsdatascience.com/a-beginners-guide-to-machine-learning-algorithms-e7615b537381>

<https://www.coursera.org/learn/machine-learning>

<https://www.analyticsvidhya.com/blog/2015/09/a-comprehensive-guide-to-machine-learning-algorithms/>

Video Links:

https://www.youtube.com/watch?v=4UJeIID_ICw

<https://www.youtube.com/watch?v=h0e2HAPTGF4>

<https://www.youtube.com/watch?v=ukzFI9rgwfU>

https://www.youtube.com/watch?v=i_LwzRVP7bg

<https://www.youtube.com/watch?v=IPM3ISFQy68>

https://www.youtube.com/watch?v=jGwO_UgTS7I

<https://www.youtube.com/watch?v=7uwa9aPbBRU>

<https://www.youtube.com/watch?v=FlpbrnbVfJk>

<https://www.youtube.com/watch?v=QpzQYDcczRQ>

<https://www.youtube.com/watch?v=7kyNYMwfYdw>

पर्यवेक्षति शिक्षण तकनीकें

सुपरवाइज्ड लर्नगि मशीन लर्नगि का एक मूलभूत पहलू है जो पूर्वानुमान या निर्णय लेने के लिए लेबल किए गए प्रशिक्षण डेटा पर निर्भर करता है। इस उप-मॉड्यूल में, हम इसकी परिभाषा, प्रमुख तकनीकों और अनुप्रयोगों का पता लगाएंगे। विभिन्न एल्गोरिदम और वास्तविक दुनिया के उदाहरणों की जांच करके, छात्रों को इस बात की जानकारी मिलेगी कि विभिन्न डोमेन में सुपरवाइज्ड लर्नगि को प्रभावी ढंग से कैसे लागू किया जा सकता है।

पर्यवेक्षति शिक्षण की परिभाषा

सुपरवाइज्ड लर्नगि एक प्रकार की मशीन लर्नगि है, जिसमें मॉडल को लेबल किए गए डेटासेट पर प्रशिक्षित किया जाता है। इस सेटअप में, प्रत्येक प्रशिक्षण उदाहरण में इनपुट डेटा (फीचर्स) और संबंधित आउटपुट (लेबल) शामिल होते हैं। इसका उद्देश्य मॉडल के लिए इनपुट और आउटपुट के बीच संबंध सीखना और नए, अनदेखे डेटा के लिए आउटपुट की भविष्यवाणी करना है। उदाहरण के लिए, स्पैम का पता लगाने के लिए सुपरवाइज्ड लर्नगि मॉडल में, ईमेल (इनपुट डेटा) को 'स्पैम' या 'स्पैम नहीं' (आउटपुट) के रूप में लेबल किया जाता है। एल्गोरिदम इन उदाहरणों से सीखता है और भविष्य के ईमेल को वर्गीकृत कर सकता है।

पर्यवेक्षति शिक्षण में प्रमुख तकनीकें

कई तकनीकें पर्यवेक्षति शिक्षण की नींव बनाती हैं। सबसे उल्लेखनीय में शामिल हैं: 1. **रैखिक प्रतिगमन**¹: एक विधि जिसका उपयोग आश्रित चर और एक या अधिक स्वतंत्र चर के बीच संबंध को मॉडल करने के लिए किया जाता है। उदाहरण के लिए, वरग फुटेज, बेडरूम की संख्या और स्थान के आधार पर घर की कीमतों की भविष्यवाणी करना। 2. **लॉजिस्टिक प्रतिगमन**²: बाइनरी वर्गीकरण समस्याओं के लिए एक सांख्यिकीय विधि। एक उदाहरण में यह अनुमान लगाना शामिल है कि कोई ग्राहक अपनी जनसांख्यिकी और पछिले व्यवहार के आधार पर कोई उत्पाद खरीदेगा या नहीं (हाँ/नहीं)। 3. **सपोर्ट वेक्टर मशीन (SVM)**³: एक एल्गोरिथम जो हाइपरप्लेन को खोजता है जो डेटासेट में विभिन्न वर्गों को सबसे अच्छी तरह से अलग करता है। उदाहरण के लिए, का उपयोग पंखुड़ियों और सीपल माप के आधार पर फूलों की विभिन्न प्रजातियों की पहचान करने में किया जा सकता है। 4. **निरणय वृक्ष**⁴: एक मॉडल

जो डेटा को वर्गीकृत करने के लिए नरिणयों के वृक्ष-जैसे ग्राफ का उपयोग करता है। उदाहरण के लिए, आय, क्रेडिट स्कोर और रोजगार की स्थिति जैसे आवेदक विशेषताओं के आधार पर ऋण स्वीकृत किया जाना चाहिए या नहीं, यह नरिधारित करने के लिए नरिणय वृक्ष का उपयोग किया जा सकता है। 5. ****रैंडम फॉरेस्ट****: एक समूह तकनीक जो पूर्वानुमानों में बेहतर सटीकता और मजबूती के लिए कई नरिणय वृक्षों का उपयोग करती है। इसे वभिनिन विशेषताओं के आधार पर ग्राहक चर्न की भविष्यवाणी करने जैसे कार्यों के लिए नयोजित किया जा सकता है। 6. ****न्यूरल नेटवर्क****: मानव भस्तषिक संरचना से प्रेरित एक मॉडल जो जटिल वर्गीकरण और प्रतगिमन कार्यों, जैसे छवि पहचान या पाठ प्रसंस्करण के लिए अत्यधिक प्रभावी है।

पर्यवेक्षति शकिषण के अनुप्रयोग

पर्यवेक्षति शकिषण के वभिनिन उद्योगों में वभिनिन अनुप्रयोग हैं। यहाँ कुछ उल्लेखनीय उपयोग-मामले दिए गए हैं: - ****सेवास्थय सेवा****: रोगी डेटा के आधार पर रोग के परिणामों की भविष्यवाणी करना। उदाहरण के लिए, चिकित्सा इतिहास और जीवनशैली कारकों के आधार पर मधुमेह संबंधी जटिलताओं की भविष्यवाणी करने के लिए पर्यवेक्षति शकिषण का उपयोग करना। - ****वर्तित****: क्रेडिट स्कोरिंग मॉडल जो उधारकर्ता द्वारा ऋण पर चुक की संभावना का मूल्यांकन करते हैं। पछिले ऋण डेटा पर मॉडल को प्रशकिषति करके, बैंक उच्च जोखिम वाले ग्राहकों की पहचान कर सकते हैं। - ****मार्केटगि****: मार्केटगि रणनीतियों को वैयक्तिकृत करने के लिए कर्ये व्यवहार और जनसांख्यिकी के आधार पर ग्राहक वभिजन। - ****धोखाधड़ी का पता लगाना****: लेबल किए गए लेनदेन से पैटर्न का विश्लेषण करके वास्तविक समय में धोखाधड़ी वाले लेनदेन की पहचान करना।

चुनौतियाँ और वचिर

इसकी प्रभावशीलता के बावजूद, पर्यवेक्षति शकिषण को नमिन चुनौतियों का सामना करना पड़ता है: - ****डेटा की गुणवत्ता****: मॉडल का प्रदर्शन लेबल किए गए डेटा की गुणवत्ता पर बहुत अधिक निर्भर करता है। खराब लेबल गलत पूर्वानुमानों को जन्म दे सकते हैं। - ****ओवरफिटिंग****: यदि कोई मॉडल बहुत जटिल है, तो यह प्रशकिषण डेटा पर अच्छा प्रदर्शन कर सकता है लेकिन अनदेखे डेटा पर खराब प्रदर्शन कर सकता है। इसे कम करने के लिए क्रॉस-वैलिडेशन और रेग्यूलराइजेशन जैसी तकनीकों का उपयोग किया जाता है। - ****क्लास में असंतुलन****: यदि डेटासेट में एक क्लास का प्रतनिधित्व काफी कम है, तो मॉडल बहुसंख्यक क्लास के प्रतपक्षपाती हो सकता है। इसे संबोधित करने के लिए रीसैपलिंग या वभिनिन मूल्यांकन मीट्रिक का उपयोग करने जैसी रणनीतियाँ आवश्यक हैं।

नषिर्क्ष

पर्यवेक्षति शकिषण तकनीकें शक्तशाली उपकरण हैं जो वभिनिन अनुप्रयोगों में पूर्वानुमानित मॉडलिंग को सक्षम बनाती हैं। वभिनिन एल्गोरिदम और उनके उपयुक्त उपयोग-मामलों को समझना छात्रों को वास्तविक दुनिया की समस्याओं में इन वधियों को प्रभावी ढंग से लागू करने में सक्षम बना सकता है। पर्यवेक्षति शकिषण में महारत हासिल करके, छात्र परिष्कृत मॉडल बना सकते हैं जो कई क्षेत्रों में नरिणय लेने की प्रक्रियाओं को संचालित करते हैं।

Reference:

<https://www.coursera.org/learn/machine-learning>

<https://www.edx.org/course/machine-learning-fundamentals>

https://scikit-learn.org/stable/supervised_learning.html

<https://towardsdatascience.com/supervised-learning-in-machine-learning-e9a2b30b6e4b>

<https://www.analyticsvidhya.com/blog/2020/10/decision-trees-in-machine-learning/>

Video Links:

https://www.youtube.com/watch?v=vvODQqb3D_8

https://www.youtube.com/watch?v=YIGEYyEM_a8

<https://www.youtube.com/watch?v=7w0hBmzkTIA>

<https://www.youtube.com/watch?v=wL3KyYurkSk>

<https://www.youtube.com/watch?v=KyGFVrcuMZM>

<https://www.youtube.com/watch?v=iGJ1XSkCyU0>

https://www.youtube.com/watch?v=O_tMgYcA8j0

<https://www.youtube.com/watch?v=sJzuNAisXHA>

https://www.youtube.com/watch?v=tJmVf88I_ys

<https://www.youtube.com/watch?v=N3zTECohLjI>

अपर्यवेक्षति शक्तिषण वधियिँ

अनसुपरवाइज्ड लरनगि मशीन लरनगि की एक श्रेणी है जो डेटा के भीतर पैटर्न या संरचनाओं की खोज पर ध्यान केंद्रित करती है जिसमें लेबल वाले आउटपुट या श्रेणियाँ नहीं होती हैं। सुपरवाइज्ड लरनगि के विपरीत, जहाँ एक मॉडल को लेबल दिए गए इनपुट-आउटपुट जोड़े पर प्रशिक्षित किया जाता है, अनसुपरवाइज्ड लरनगि लेबल रहित डेटा में छिपे हुए संबंधों को खोजता है। यह उप-मॉड्यूल अनसुपरवाइज्ड लरनगि की प्रमुख अवधारणाओं, पद्धतियों और अनुप्रयोगों में गहराई से उतरता है, जिससे छात्रों को मशीन लरनगि के क्षेत्र में इसके महत्व और कार्यक्षमता की व्यापक समझ मिलती है।

1. अपर्यवेक्षति शक्तिषण की परिभाषा

अनसुपरवाइज्ड लरनगि मशीन लरनगि की एक शाखा है, जहाँ एल्गोरिदम का उपयोग लेबल रहित डेटासेट का विश्लेषण और क्लस्टर करने के लिए किया जाता है। सिस्टम डेटा से स्वतंत्र रूप से पैटर्न और संरचना सीखने की कोशिश करता है, जिससे डेटासेट के भीतर आंतरिक संबंधों और समूहों की पहचान की जा सके। चूँकि डेटा लेबल नहीं किया जाता है, इसलिए प्रक्रिया में आमतौर पर विशिष्ट आउटपुट के माध्यम से सीखने के बजाय उदाहरण द्वारा सीखना शामिल होता है।

2. अपर्यवेक्षति शक्तिषण की प्रमुख वधियिँ

कुई प्रमुख वधियिँ अपर्यवेक्षति शक्तिषण की विशेषता बताती हैं, जिनमें शामिल हैं: - ****क्लस्टरगि****: इसमें वस्तुओं के एक समूह को इस तरह से समूहीकृत करना शामिल है कि एक ही समूह (या क्लस्टर) में वस्तुएँ अन्य समूहों की तुलना में एक-दूसरे से अधिक समान हों। उदाहरणों में -मीनस क्लस्टरगि, पदानुक्रमित क्लस्टरगि और शामिल हैं। - ****आयाम न्यूनीकरण****: (प्रसिपिल कंपोनेंट एनालिसिस) और - (डिस्ट्रिब्यूटेड स्टोचैस्टिक नेबर एम्बेडिंग) जैसी तकनीकों का उपयोग विचाराधीन यादृच्छिक चरों की संख्या को कम करने के लिए किया जाता है, जो सबसे महत्वपूर्ण विशेषताओं पर ध्यान केंद्रित करता है। - ****एसोसिएशन लरनगि****: इस वधि में बड़े डेटाबेस में चरों के बीच दलित्सु संबंधों की खोज करना शामिल है। एक विशिष्ट उदाहरण मार्केट बास्केट विश्लेषण है, जहाँ यह पता चल सकता है कि जो ग्राहक ब्रेड खरीदते हैं वे अक्सर मक्खन भी खरीदते हैं।

3. व्यावहारिक उदाहरण

कुई व्यावहारिक अनुप्रयोग अपर्यवेक्षति शक्तिषण वधियों को चिह्नित कर सकते हैं: - ****ग्राहक विभाजन****: व्यवसाय ग्राहक डेटा पर क्लस्टरगि एल्गोरिदम का उपयोग करके क्रय व्यवहार के आधार पर ग्राहकों को अलग-अलग खंडों की पहचान कर सकते हैं, जो लक्षित विपणन रणनीतियों में मदद करता है। -

****वसिंगतिका पता लगाना****: साइबर सुरक्षा में, अप्रशिक्षित शक्ति नेटवर्क ट्रैफिक में असामान्य पैटर्न की पहचान कर सकता है जो संभावित सुरक्षा उल्लंघन का संकेत दे सकता है। - ****छवि संपीड़न****: आयाम में कमी तकनीकों को लागू करके, बड़ी छवि फाइलों को आवश्यक विवरणों को संरक्षित करते हुए संपीड़ित किया जा सकता है, जिससे तेज़ अपलोड और कम संग्रहण स्थान की सुविधा मिलती है।

4. वास्तविक दुनिया के अनुप्रयोग

वभिन्न क्षेत्रों में अप्रशिक्षित शक्ति प्रचलित है: - ****स्वास्थ्य सेवा****: क्लस्टरिंग आनुवंशिक जानकारी के आधार पर रोगियों के बीच रोग पैटर्न की पहचान करने में मदद कर सकता है। - ****वैतित****: असामान्य लेनदेन को पहचानने के लिए धोखाधड़ी का पता लगाने वाली प्रणालियों में वसिंगतिका पता लगाने वाले एल्गोरिदम का व्यापक रूप से उपयोग किया जाता है जो स्थापित पैटर्न से अलग होते हैं। - ****सोशल मीडिया विश्लेषण****: ब्रांड सोशल मीडिया प्लेटफॉर्म से प्राप्त असंरचित डेटा से ग्राहक भावना और जुड़ाव के रुझानों का विश्लेषण करने के लिए अप्रशिक्षित शक्ति का उपयोग करते हैं।

5. अपर्यवेक्षित शक्ति में चुनौतियाँ

अपनी क्षमताओं के बावजूद, अप्रशिक्षित शक्ति कई चुनौतियाँ प्रस्तुत करता है: - ****सफलता के लिए कोई स्पष्ट मीट्रिक नहीं****: पर्यवेक्षित शक्ति के विपरीत, यह मापने का कोई सीधा तरीका नहीं है कि मॉडल कितना अच्छा प्रदर्शन कर रहा है क्योंकि कोई लेबल किए गए आउटपुट नहीं है। - ****एल्गोरिदम का चयन****: सही क्लस्टरिंग या आयाम में कमी करने वाले एल्गोरिदम का चयन करना जटिल और संदर्भ-निर्भर हो सकता है, जिसके लिए डोमेन ज्ञान की आवश्यकता होती है। - ****डेटा गुणवत्ता****: एल्गोरिदम की प्रभावशीलता उपयोग किए गए डेटा की गुणवत्ता पर काफी हद तक निर्भर करती है, जिसमें शोर और अप्रासंगिक विशेषताएँ शामिल हैं।

6. सारांश और निष्कर्ष

मशीन लर्निंग में अनसुपरवाइज्ड लर्निंग एक शक्तिशाली प्रतिभा है जो लेबल किए गए आउटपुट के बिना छिपे हुए पैटर्न को उजागर करता है। यह क्लस्टरिंग, आयाम में कमी और एसोसिएशन विधियों पर बहुत अधिक निर्भर करता है, जो वभिन्न उद्योगों में कई वास्तविक दुनिया के अनुप्रयोग प्रदान करता है। हालाँकि, स्पष्ट प्रदर्शन मीट्रिक की कमी और डेटा गुणवत्ता के प्रतिनिधित्व जैसी चुनौतियों को इसकी पूरी क्षमता का लाभ उठाने के लिए सावधानीपूर्वक प्रबंधित किया जाना चाहिए।

Reference:

<https://towardsdatascience.com/understanding-unsupervised-learning-5529752327f3>

<https://www.coursera.org/learn/machine-learning/lecture/sCaU2/the-basics-of-unsupervised-learning>

<https://www.analyticsvidhya.com/blog/2021/06/a-beginners-guide-to-unsupervised-learning-in-machine-learning/>

<https://www.kdnuggets.com/2020/04/introduction-unsupervised-learning.html>

Video Links:

<https://www.youtube.com/watch?v=8dqdDEyzkFA>

https://www.youtube.com/watch?v=vvODQqb3D_8

<https://www.youtube.com/watch?v=5yeJ03crTrI>

<https://www.youtube.com/watch?v=qDbpYUbf3e0>

<https://www.youtube.com/watch?v=D6gtZrsYi6c>

<https://www.youtube.com/watch?v=ECHY1tD2b6Q>

<https://www.youtube.com/watch?v=3fsy2oheRdg>

https://www.youtube.com/watch?v=W01tIRP_Rqs

<https://www.youtube.com/watch?v=3TbU9HI8lo4>

https://www.youtube.com/watch?v=YIGeqyEM_a8

एल्गोरिदम के लिए मूल्यांकन मेट्रिक्स

मशीन लर्निंग में, एल्गोरिदम के प्रदर्शन का आकलन करने के लिए मूल्यांकन मीट्रिक महत्वपूर्ण हैं। वे मात्रात्मक उपाय प्रदान करते हैं जो हमें यह समझने में मदद करते हैं कि हमारे मॉडल दिए गए डेटासेट के आधार पर कतिनी अच्छी तरह से भविष्यवाणियाँ करते हैं। चाहे आप वर्गीकरण, प्रतगमन या क्लस्टरिंग समस्या पर काम कर रहे हों, सही मीट्रिक जानने से बेहतर मॉडल चयन, अनुकूलन और सत्यापन संभव होता है।

1. वर्गीकरण मेट्रिक्स

वर्गीकरण मेट्रिक्स का उपयोग उन एल्गोरिदम का मूल्यांकन करने के लिए किया जाता है जो असतत लेबल की भविष्यवाणी करते हैं। सबसे आम मेट्रिक्स में शामिल हैं: - ****सटीकता****: कुल भविष्यवाणियों के लिए सही भविष्यवाणियों का अनुपात। सरल होते हुए भी, यह भ्रामक हो सकता है, खासकर असंतुलित वर्गों के साथ। - ****उदाहरण***: यदि कोई मॉडल 100 में से 90 बार सही भविष्यवाणी करता है, तो सटीकता 90% होगी। - ****परिशुद्धता****: सचचे सकारात्मक (टीपी) का सचचे सकारात्मक और झूठे सकारात्मक (एफपी) के योग से अनुपात। यह मूल्यांकन करने में मदद करता है कि अनुमानित सकारात्मक मामलों में से कतिने वास्तव में सकारात्मक थे। - ****उदाहरण***: यदि कोई मॉडल 10 सकारात्मक मामलों (8 सही) की भविष्यवाणी करता है, तो परिशुद्धता = $8 / (8 + 2) = 0.8$ । - ****रिक्ल (संवेदनशीलता)****: सचचे सकारात्मक और झूठे नकारात्मक (एफएन) के योग से सचचे सकारात्मक का अनुपात। यह दर्शाता है कि मॉडल सकारात्मक उदाहरणों को कतिनी अच्छी तरह से कैपचर करता है। - ****परिशुद्धता****: सचचे सकारात्मक और झूठे नकारात्मक (एफएन) के योग से सचचे सकारात्मक का अनुपात। यह दर्शाता है कि मॉडल सकारात्मक उदाहरणों को कतिनी अच्छी तरह से कैपचर करता है। - ****परिशुद्धता****: सचचे सकारात्मक और झूठे नकारात्मक (एफएन) के योग से सचचे सकारात्मक का अनुपात। - ***उदाहरण***: यदि 10 वास्तविक सकारात्मक मामले हैं और मॉडल 8 को कैपचर करता है, तो रिक्ल = $8 / (8 + 2) = 0.8$ । - ****1 स्कोर****: परिशुद्धता और रिक्ल का हार्मोनिक माध्य। यह दो मेट्रिक्स को संतुलित करता है, विशेष रूप से तब उपयोगी होता है जब आपको परिशुद्धता और रिक्ल के बीच ट्रेड-ऑफ को व्यक्त करने की आवश्यकता होती है। - ***उदाहरण***: यदि परिशुद्धता = 0.8 और रिक्ल = 0.8, तो 1 स्कोर = $2 * (0.8 * 0.8) / (0.8 + 0.8) = 0.8$ ।

2. रजिशन मेट्रिक्स

रजिशन मेट्रिक्स उन एल्गोरिदम के प्रदर्शन का आकलन करते हैं जो नरितर परिणामों की भविष्यवाणी करते हैं। मुख्य मेट्रिक्स में शामिल हैं: - ****माध्य नरिपेक्ष त्रुटि (****: पूर्वानुमानित और वास्तविक मूल्यों के बीच नरिपेक्ष अंतर का औसत। यह पूर्वानुमानित मूल्य के समान इकाइयों में औसत त्रुटि प्रदान करता है। - ***उदाहरण***: यदि पूर्वानुमानित मान [2, 3, 5] हैं और वास्तविक मान [2, 7, 4] हैं, तो $= (|2-2| + |3-7| + |5-4|) / 3 = 1.33$ । - ****माध्य वर्ग त्रुटि (****: पूर्वानुमानित और वास्तविक मूल्यों के बीच वर्ग अंतर का औसत। यह वर्गीकरण के कारण बड़ी त्रुटियों पर जोर देता है। - ***उदाहरण***: ऊपर बताए गए समान पूर्वानुमानों का उपयोग करते हुए, $= ((2-2)^2 + (3-7)^2 + (5-4)^2) / 3 = (0 + 16 + 1) / 3 = 5.67$ । - ****स्क्वायरड****: वास्तविक मानों के औसत की तुलना में मॉडल द्वारा बताए गए वचरण के अनुपात को दर्शाता है। 1 के करीब के मान एक अच्छे फिट का संकेत देते हैं। - ***उदाहरण***: यदि वास्तविक मानों में कुल भिन्नता 100 है, और मॉडल इसका 80 बताता है, तो -स्क्वायरड = $80 / 100 = 0.8$ ।

3. क्लस्टरिंग मेट्रिक्स

क्लस्टरिंग मेट्रिक्स उन एल्गोरिदम के प्रदर्शन का मूल्यांकन करते हैं जो बिना किसी पूर्वनिर्धारित लेबल के समानता के आधार पर डेटा बिंदुओं को समूहीकृत करते हैं। सामान्य मेट्रिक्स में शामिल हैं: - ****सिल्यूट स्कोर****: मापता है कि कोई ऑब्जेक्ट अन्य क्लस्टर की तुलना में अपने क्लस्टर से कतिना समान है। यह -1 से 1 तक होता है, जिसमें उच्च मान बेहतर प्रभावित क्लस्टर को इंगित करते हैं। - ***उदाहरण***: 0.5 का सिल्यूट स्कोर बताता है कि ऑब्जेक्ट अपने क्लस्टर से अच्छी तरह मेल खाता है, लेकिन पड़ोसी क्लस्टर के करीब हो सकता है। - ****डेविस-बोल्डन इंडेक्स****: क्लस्टर के भीतर बखिराव और क्लस्टर के बीच अलगाव का अनुपात, जहां कम मान बेहतर क्लस्टरिंग को इंगित करते हैं। - ***उदाहरण***: यदि क्लस्टरिंग एल्गोरिदम का परिणाम 0.7 का डेविस-बोल्डन इंडेक्स है, तो यह सुझाव देता है कि क्लस्टर कॉम्पैक्ट और अच्छी तरह से अलग हैं।

4. मूल्यांकन मेट्रिक्स का महत्व

सही मूल्यांकन मीट्रिक को समझना और उसका उपयोग करना नमिन के लिए आवश्यक है: - ****मॉडल चयन****: मूल्यांकन परिणामों के आधार पर किसी समस्या को हल करने के लिए उपयुक्त एल्गोरिदम चुनना। - ****पैरामीटर ट्युनिंग****: मॉडल पैरामीटर को अनुकूलित करने से चयनित मीट्रिक के अनुसार प्रदर्शन में सुधार होता है। - ****सत्यापन****: यह सुनिश्चित करना कि मॉडल अदृश्य डेटा के लिए अच्छी तरह से सामान्यीकृत हो, जिससे ओवरफिटिंग को रोका जा सके।

5. व्यावहारिक अनुप्रयोग और केस अध्ययन

मूल्यांकन मीट्रिक्स का उपयोग विभिन्न क्षेत्रों में व्यापक रूप से किया जाता है: - ****स्वास्थ्य सेवा****: चिकित्सा निदान में, सटीकता और स्मरण, रोगों की भविष्यवाणी करने वाले एल्गोरिदम को आकलन करने के लिए महत्वपूर्ण है, यह सुनिश्चित करते हुए कि रोगियों को गलत निदान के बिना आवश्यक हस्तक्षेप प्राप्त हो। - ****वित्त****: धोखाधड़ी का पता लगाने में, झूठे सकारात्मक परिणामों को कम करने के लिए उच्च प्रशुद्धता महत्वपूर्ण है, जो वैध लेनदेन को बाधित कर सकते हैं। - ****ई-कॉमर्स****: अनुशंसा प्रणालियों में, विविध उत्पादों की अनुशंसा करने और उच्च प्रासंगिकता बनाए रखने के बीच संतुलन बनाने के लिए 1-स्कोर जैसे मीट्रिक्स के माध्यम से उपयोगकर्ता वरीयताओं को समझना बढ़ाया जाता है।

Reference:

https://scikit-learn.org/stable/modules/model_evaluation.html

<https://towardsdatascience.com/evaluation-metrics-in-machine-learning-51f80b1957d1>

<https://www.analyticsvidhya.com/blog/2020/11/evaluation-metrics-for-classification-in-machine-learning/>

<https://machinelearningmastery.com/regression-metrics-for-machine-learning/>

Video Links:

<https://www.youtube.com/watch?v=wpQiEHYkByS>

https://www.youtube.com/watch?v=47kv_jxab4o

https://www.youtube.com/watch?v=FeKSQy5t_TI

<https://www.youtube.com/watch?v=nQxfuGKUQiU>

<https://www.youtube.com/watch?v=pyl6fO4C7h4>

https://www.youtube.com/watch?v=OFXO_cUPY_4

<https://www.youtube.com/watch?v=urtXN7FJre4>

<https://www.youtube.com/watch?v=aWAnNHXIKww>

https://www.youtube.com/watch?v=lt1YxJ_8Jzs

<https://www.youtube.com/watch?v=F6gpHfx3BCs>

फीचर चयन और इंजीनियरिंग

फीचर चयन और इंजीनियरिंग मशीन लर्निंग वर्कफ़्लो में महत्वपूर्ण चरण है जो सीधे पूर्वानुमानित मॉडल के प्रदर्शन को प्रभावित करते हैं। फीचर चयन में आपके डेटासेट से सबसे प्रासंगिक फीचर (चर) की पहचान करना शामिल है जो मॉडल की सटीकता में योगदान करते हैं, जबकि फीचर इंजीनियरिंग में मॉडल की सीखने की क्षमता को बेहतर बनाने के लिए मौजूदा डेटा के आधार पर नए फीचर बनाना शामिल है। इन प्रक्रियाओं को समझने से अधिक कुशल, व्याख्या करने योग्य और उच्च प्रदर्शन वाले मॉडल बनाने में मदद मिलती है।

फीचर चयन की परिभाषा

फीचर चयन मॉडल निर्माण में उपयोग के लिए प्रासंगिक सुविधाओं के एक उपसमूह का चयन करने की प्रक्रिया है। इसका उद्देश्य डेटासेट की आयामीता को कम करना, मॉडल प्रदर्शन को बढ़ाना और ओवरफिटिंग को कम करना है। इसमें प्रत्येक विशेषता की प्रासंगिकता का मूल्यांकन करना और यह तय करना शामिल है कि कौन सी विशेषताओं को बनाए रखा जाना चाहिए और कौन सी त्याग दिया जाना चाहिए। फीचर चयन की तकनीकों में फिल्टर विधियाँ (सांख्यिकीय उपायों का उपयोग करना), रैपर विधियाँ (फीचर उपसमूहों का आकलन करने के लिए एक पूर्वानुमानित मॉडल का उपयोग करना) और एम्बेडेड विधियाँ (मॉडल प्रशिक्षण प्रक्रिया के भीतर फीचर चयन को एकीकृत करना) शामिल हैं।

फीचर इंजीनियरिंग की परिभाषा

फीचर इंजीनियरिंग का तात्पर्य अंतरनिहित समस्या को पूर्वानुमानित मॉडल में बेहतर ढंग से प्रस्तुत करने के लिए नई सुविधाओं के निर्माण या मौजूदा सुविधाओं के परिवर्तन से है। इस प्रक्रिया में सुविधाओं को स्केल करना, श्रेणीबद्ध चर को एनकोड करना, अंतःक्रियात्मक शब्द बनाना या बहुपद सुविधाएँ उत्पन्न करना शामिल हो सकता है। प्रभावी फीचर इंजीनियरिंग मॉडल की सटीकता और मजबूती को महत्वपूर्ण रूप से बढ़ा सकती है। सार्थक सुविधाएँ प्राप्त करने के लिए डोमेन ज्ञान और रचनात्मकता की आवश्यकता होती है।

फीचर चयन में सामान्य तकनीकें

[{'तकनीक': 'फिल्टर विधियाँ', 'विवरण': 'ये विधियाँ सांख्यिकीय परीक्षणों और सहसंबंध गुणांक, ची-स्क्वायर परीक्षण और पारस्परिक जानकारी जैसे मेट्रिक्स का उपयोग करके सुविधाओं के महत्व का मूल्यांकन करती हैं। वे मॉडल-अज्ञेयवादी हैं और कंप्यूटेशनल रूप से कुशल हो सकते हैं।'}, {'तकनीक': 'रैपर विधियाँ', 'विवरण': 'रैपर विधियाँ सुविधाओं के संयोजनों का मूल्यांकन करने के लिए एक पूर्वानुमान मॉडल का उपयोग करती हैं। पुनरावर्ती सुविधा उन्मूलन () और फॉरवर्ड सिलेक्शन जैसी तकनीकें इस श्रेणी में आती हैं। वे अधिक सटीक होती हैं लेकिन कंप्यूटेशनल रूप से महंगी होती हैं।'}, {'तकनीक': 'एम्बेडेड विधियाँ', 'विवरण': 'इन विधियों में मॉडल प्रशिक्षण प्रक्रिया के भाग के रूप में सुविधा चयन शामिल है। लैसो या रिएनिय वृक्ष जैसे एल्गोरिदम स्वाभाविक रूप से मॉडल को फिट करते समय सुविधा चयन करते हैं।'}]

फीचर इंजीनियरिंग में सामान्य तकनीकें

[{'तकनीक': 'सामान्यीकरण/मानकीकरण', 'विवरण': 'फीचर्स को एक सामान्य पैमाने पर परिवर्तित करने से डेटा के पैमाने के प्रति संवेदनशील एल्गोरिदम में मदद मिल सकती है, जैसे कि ग्रेडिएंट डिसेंट-आधारित एल्गोरिदम।'}, {'तकनीक': 'श्रेणीबद्ध चर को एनकोड करना', 'विवरण': 'वन-हॉट एनकोडिंग या लेबल

एनकोडिंग जैसी तकनीकों श्रेणीबद्ध विशेषताओं को मॉडल प्रशिक्षण के लिए उपयुक्त संख्यात्मक प्रारूपों में परिवर्तित करती हैं।', {'तकनीक': 'इंटरैक्शन फीचर्स बनाना'}, 'विवरण': 'मौजूदा सुविधाओं के बीच इंटरैक्शन के माध्यम से नई सुविधाएँ उत्पन्न करना उन संबंधों को पकड़ सकता है जो मॉडल में मूल्य जोड़ते हैं।', {'तकनीक': 'आयाम में कमी'}, 'विवरण': 'प्रसिपिल कंपोनेंट एनालिसिस (पीसीए) जैसी विधियाँ डेटा में भिन्नता को संरक्षित करते हुए सुविधाओं की संख्या को कम करती हैं।']}]

काल्पनिक परिदृश्य: घर की कीमतों का पूर्वानुमान

घर की कीमतों का अनुमान लगाने के लिए एक डेटासेट पर विचार करें जिसमें 'आकार', 'स्थान', 'बेडरूम की संख्या' और 'संपत्ति की आयु' जैसी विशेषताएँ शामिल हों। सबसे प्रासंगिक विशेषताओं का चयन करने के लिए, आप उपयोग कर सकते हैं: 1. फ़िल्टर विधियाँ: लक्ष्य चर (घर की कीमत) के साथ प्रत्येक विशेषता के सहसंबंध की गणना करें और एक नशिचति सीमा से नीचे की विशेषताओं को हटा दें। 2. रैपर विधियाँ: सुविधाओं के किस संयोजन से सबसे अच्छा मॉडल प्रदर्शन होता है, इसका आकलन करने के लिए एक रेखीय प्रतगिमान मॉडल का उपयोग करें। फीचर इंजीनियरिंग के लिए, आप यह कर सकते हैं: 1. 'आकार * स्थान गुणवत्ता' जैसी इंटरैक्शन सुविधाएँ बनाएँ ताकि यह पता लगाया जा सके कि विभिन्न स्थानों पर आकार कीमत को कैसे प्रभावित करता है। 2. यह सुनिश्चित करने के लिए कि पुरानी संपत्तियों का उचित रूप से प्रतिनिधित्व किया जाता है, 'संपत्ति की आयु' सुविधा को सामान्य करें।

फीचर चयन और इंजीनियरिंग के वास्तविक-विश्व अनुप्रयोग

वर्तित में, फीचर चयन का उपयोग उन प्रमुख वृत्तीय संकेतकों की पहचान करने के लिए किया जाता है जो स्टॉक की कीमतों की भविष्यवाणी करते हैं। स्वास्थ्य सेवा में, रोगी डेटा से नई सुविधाएँ प्राप्त करना रोग पूर्वानुमान मॉडल में सहायता कर सकता है। फीचर इंजीनियरिंग तकनीकों को अनुशंसा प्रणालियों में भी देखा जा सकता है जहाँ उपयोगकर्ता की प्राथमिकताओं को उपयोगकर्ता अनुभव को बढ़ाने के लिए प्रभावी ढंग से एनकोड किया जाता है।

Reference:

<https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-24e7d922f8f6>

<https://towardsdatascience.com/feature-engineering-for-machine-learning-what-it-is-and-how-to-do-it-5c7c8c1f2f2e>

https://scikit-learn.org/stable/modules/feature_selection.html

<https://www.analyticsvidhya.com/blog/2018/03/feature-engineering-definitive-guide/>

Video Links:

<https://www.youtube.com/watch?v=uWnD04nJbDQ>

<https://www.youtube.com/watch?v=WEIBhXr9B7c>

<https://www.youtube.com/watch?v=pYVScuY-GPk>

<https://www.youtube.com/watch?v=mHrjhif2Cgw>

<https://www.youtube.com/watch?v=nzxbfQcQgk4>

<https://www.youtube.com/watch?v=blvmNWbcPD0>

<https://www.youtube.com/watch?v=fQmHNN2-KQU>

https://www.youtube.com/watch?v=vsKNxbP8R_8

<https://www.youtube.com/watch?v=ZckQtif4v1o>

<https://www.youtube.com/watch?v=4-4pKPv9IJ4>

एल्गोरिदम अनुकूलन तकनीकें

एल्गोरिदम अनुकूलन तकनीकें मशीन लर्नगि में महत्वपूर्ण हैं क्योंकि वे पूर्वानुमानित मॉडल प्राप्त करने में एल्गोरिदम के प्रदर्शन, दक्षता और सटीकता को सीधे प्रभावित करती हैं। यह उप-मॉड्यूल बेहतर संसाधन उपयोग, कम कम्प्यूटेशनल समय और बेहतर पूर्वानुमान कक्षमताओं के लिए एल्गोरिदम की अनुकूलति करने के लिए उपयोग की जाने वाली विभिन्न रणनीतियों और विधियों की खोज करता है। इन तकनीकों की समझना किसी भी महत्वाकांक्षी मशीन लर्नगि व्यवसायी के लिए आवश्यक है ताकि यह सुनिश्चित हो सके कि उनके मॉडल वास्तविक दुनिया के अनुप्रयोगों में उत्कृष्ट प्रदर्शन करते हैं।

1. एल्गोरिदम अनुकूलन का परिचय

एल्गोरिदम अनुकूलन में एल्गोरिदम के प्रदर्शन की दक्षता में सुधार करना शामिल है - संक्षेप में, संसाधन खपत को कम करते हुए किसी कार्य को पूरा करने का सबसे अच्छा संभव तरीका खोजना। इसमें समय जटिलता, स्थान जटिलता या दोनों शामिल हो सकते हैं। सामान्य अनुकूलन लक्ष्यों में कम्प्यूटेशनल समय को कम करना, सटीकता को बढ़ाना और एल्गोरिदम की मापनीयता को बढ़ाना शामिल है।

2. मशीन लर्नगि में अनुकूलन का महत्व

मशीन लर्नगि के क्षेत्र में, जहाँ बड़े डेटासेट अक्सर होते हैं, अनुकूलन आवश्यक हो जाता है। अनुकूलन के बिना, प्रशिक्षण का समय अव्यवहारिक रूप से लेबा हो सकता है, और मॉडल वास्तविक समय की स्थितियों में तैनात करने के लिए बहुत जटिल हो सकते हैं। अनुकूलन तकनीकें सुनिश्चित करती हैं कि मॉडल तेजी से भविष्यवाणियाँ करते हुए डेटा से प्रभावी ढंग से सीख सके, जो वित्त, स्वास्थ्य सेवा और ऑनलाइन सेवाओं जैसे उद्योगों में विशेष रूप से महत्वपूर्ण है।

3. सामान्य अनुकूलन तकनीकें

[{'': '3.1 ग्रेडिएंट डिसेंस और इसके वेरिएंट', '': 'ग्रेडिएंट डिसेंस किसी फंक्शन को न्यूनतम करने के लिए सबसे आम अनुकूलन तकनीकों में से एक है। इसमें लॉस फंक्शन के ग्रेडिएंट (या ढलान) की गणना करना और विपरीत दिशा में मापदंडों को अपडेट करना शामिल है। वेरिएंट में स्टोचैस्टिक ग्रेडिएंट डिसेंस (), मनी-बैच ग्रेडिएंट डिसेंस और एडम ऑप्टिमाइजर शामिल हैं, जो सीखने की दरों को अनुकूल रूप से समायोजित करते हैं। उदाहरण के लिए, न्यूरल नेटवर्क को प्रशिक्षित करने में, एडम ऑप्टिमाइजर अपने अनुकूली दृष्टिकोण के कारण मानक ग्रेडिएंट डिसेंस की तुलना में तेजी से अभिसरित हो सकता है।'}, {'': '3.2 रेग्यूलराइजेशन तकनीकें', '': 'एल1 और एल2 रेग्यूलराइजेशन जैसे रेग्यूलराइजेशन विधियों का उपयोग लॉस फंक्शन में पेनल्टी टर्म जोड़कर ओवरफिटिंग को रोकने के लिए किया जाता है। यह सुनिश्चित करता है कि मॉडल नए डेटा पर अच्छी तरह से सामान्यीकृत हो। उदाहरण के लिए, लैससो रीग्रेशन (1 नियमितीकरण) वास्तव में कुछ गुणों को शून्य तक कम कर सकता है, ओवरफिटिंग को रोकते हुए प्रभावी रूप से फीचर चयन कर सकता है।'}, {'': '3.3 हाइपरपैरामीटर ट्युनिंग', 'सामगरी': 'हाइपरपैरामीटर ट्युनिंग में मॉडल सीखने की प्रक्रिया को नियंत्रित करने वाले मापदंडों की ठीक करना शामिल है, जैसे सीखने की दर, एनसेबल विधियों में पेडों की संख्या या नरिणय पेडों में पेडों की गहराई। ग्रडि सर्च, रैंडम सर्च और बायेसियन ऑप्टिमाइजेशन जैसी तकनीकों को इष्टतम सेटिंग्स खोजने के लिए लागू किया जा सकता है। उदाहरण के लिए, कर्नेल फंक्शन का सबसे अच्छा संयोजन खोजने के लिए ग्रडि सर्च का उपयोग करना मॉडल के प्रदर्शन को काफी हद तक बढ़ा सकता है।'}, {'': '3.4 एनसेबल विधियाँ', 'सामगरी': 'बैगिंग (बूटस्ट्रैप एग्रीगेटिंग) और बूस्टिंग जैसी एनसेबल विधियाँ समग्र प्रदर्शन को बेहतर बनाने के लिए कई मॉडलों को जोड़ती हैं। उदाहरण के लिए, रैंडम फॉरेस्ट (एक बैगिंग विधि) विभिन्न नरिणय वृक्षों के परिणामों का औसत निकालकर ओवरफिटिंग को कम करता है, जबकि बूस्टिंग (उदाहरण के लिए, एडाबूस्ट) पूर्वानुमान सटीकता में सुधार करने के लिए गलत वर्गीकृत डेटा बटुओं पर क्रमिक रूप से ध्यान केंद्रित करता है।'}, {'': '3.5 आयाम में कमी', 'सामगरी': 'प्रसिपिल कंपोनेंट एनालिसिस () और -डिस्ट्रिब्यूटेड स्टोचैस्टिक नेबर एम्बेडिंग (-) जैसी तकनीकें

महत्वपूर्ण जानकारी को संरक्षित करते हुए सुविधाओं की संख्या को कम करती है। यह न केवल प्रसंस्करण को गति देता है बल्कि मॉडल की सामान्यीकरण करने की क्षमता में भी सुधार करता है। उदाहरण के लिए, 100 सुविधाओं वाले डेटासेट पर लागू करने से इसे 10 मुख्य घटकों तक कम करने में मदद मिल सकती है, जिससे आगे की प्रक्रिया को सुव्यवस्थित किया जा सकता है।"]}

4. वास्तविक दुनिया के अनुप्रयोग

एल्गोरिदम अनुकूलन तकनीकें वभिन्न डोमेन में लागू होती हैं। वित्त में, कुशल एल्गोरिदम मिलीसेकंड में लेनदेन की प्रक्रिया करते हैं। स्वास्थ्य सेवा में, अनुकूलित मॉडल रोगी के नदान की भविष्यवाणी तेजी से और अधिक सटीक रूप से करते हैं, जिससे स्वास्थ्य सेवा के परिणाम बेहतर होते हैं। ई-कॉमर्स में, अनुशंसा प्रणाली उपयोगकर्ताओं को वास्तविक समय के सुझाव प्रदान करने के लिए अनुकूलन से लाभान्वित होती है, जिससे रूपांतरण दर बढ़ती है।

5. चुनौतियाँ और भविष्य की दिशाएँ

जैसे-जैसे मॉडल और डेटासेट बड़े होते जाते हैं, एल्गोरिदम को अनुकूलित करना अनूठी चुनौतियों का सामना करता है। चल रहे शोध हाइपरपैरामीटर ट्यूनिंग को स्वचालित करने, मॉडल जटिलता और व्याख्या के बीच व्यापार-नापसंद को समझने और अधिक कुशल एल्गोरिदम विकसित करने पर केंद्रित हैं जो कम डेटा से सीख सकते हैं। न्यूरल आर्किटेक्चर सर्च () जैसी तकनीकें न्यूरल नेटवर्क के डिज़ाइन को स्वचालित करने के लिए उभर रही हैं, जो भविष्य के अनुकूलन प्रयासों में वादा दिखाती हैं।

Reference:

<https://machinelearningmastery.com/gradient-descent-for-machine-learning/>

<https://towardsdatascience.com/the-importance-of-regularization-in-machine-learning-5c44f92914a8>

<https://www.coursera.org/learn/machine-learning-tuning>

<https://towardsdatascience.com/hyperparameter-tuning-in-machine-learning-using-grid-search-and-random-search-79476e92a452>

<https://www.analyticsvidhya.com/blog/2018/09/boosting-versus-bagging/>

Video Links:

<https://www.youtube.com/watch?v=8BQm4BVpR20>

<https://www.youtube.com/watch?v=QlcztmGaWkc>

<https://www.youtube.com/watch?v=Uy4agEMAGbM>

<https://www.youtube.com/watch?v=s-6rV1nMw0w>

<https://www.youtube.com/watch?v=Q2dewZweAtU>

<https://www.youtube.com/watch?v=fKfDeHNn3Bo>

https://www.youtube.com/watch?v=84HOL_EiJ4M

<https://www.youtube.com/watch?v=dNNZxV0RQsw>

<https://www.youtube.com/watch?v=T4iJ10TAIMg>

<https://www.youtube.com/watch?v=C96P0wYUqVk>