

G検定

出題傾向の解説

学習法・試験対策のアドバイス



アガルートアカデミー・データサイエンス講座担当講師の紹介



ヤン ジャクリン

北京生まれ・米国籍・日本育ち（茨城県民8年, 東京都民15年）

日本語・英語・中国語のトライリンガル

私の好きなものは：

言語（日・英・中 80%, スペイン語5%, 仏語5%, [Python 10%](#)）

コミュニケーション全般、文章の読み書き、運動

【経歴】

2010.3 東京大学理学部卒業

2015.3 東京大学理学系研究科 物理学専攻 博士課程修了（理学博士）

～2017.6 高エネルギー加速器研究機構（茨城県つくば）
素粒子原子核研究所 博士研究員

- ・主要研究：加速器・ナノビーム・検出器・ヒッグス粒子・暗黒物質
- ・各地大学で物理学の非常勤講師、産学連携分野の広報・通訳



2017.7～ 株式会社GRI（現職）
データ分析官・講師

最近の主な活動：

- ・マーケティング分野の分析、データ整形・可視化分析
- ・データサイエンス分野の講座、法人研修を提供
- ・データサイエンス人材育成カリキュラムの設計



G検定の概要

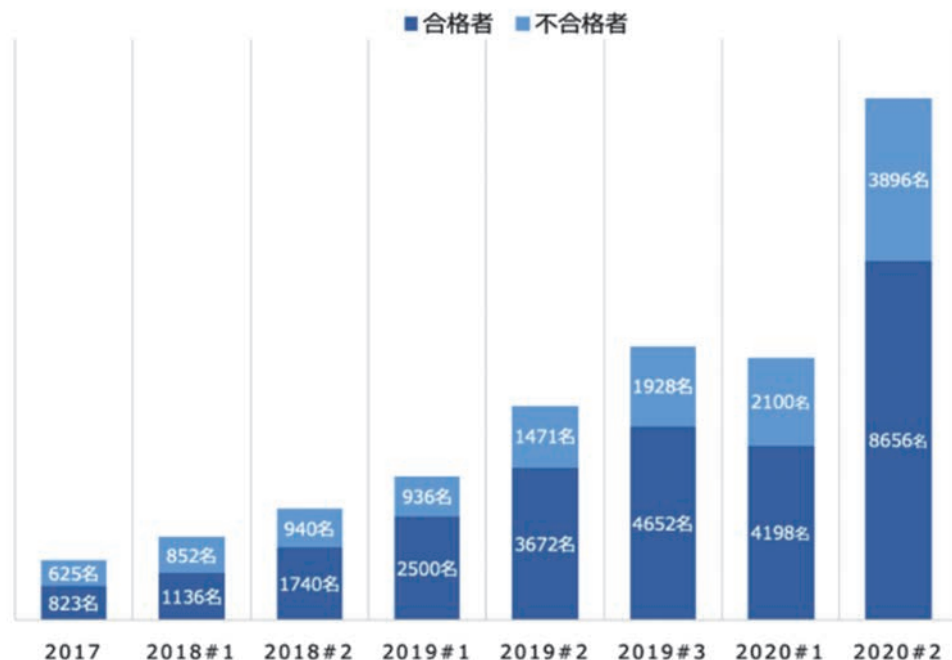
G検定（ジェネラリスト検定）の概要

- 一般社団法人日本ディープラーニング協会（JDLA）が実施
<https://www.jdla.org/business/certificate/>
- 目的
ディープラーニングを事業に活かすための知識を検定する
- 試験概要：
 - **多肢選択式の知識問題 220問程度**（シラバスより出題）
 - **人工知能やディープラーニングに関する幅広い内容が出題される**
- 試験時間：120分
- 試験形態：オンライン実施（自宅受験可）
- 受験料：一般 12,000円（税抜） 学生 5,000円（税抜）
- 試験日：年3回、日程確認と申込は <https://www.jdla-exam.org/d/>

- 2017年12月に第1回目が実施された、比較的新しい検定試験
- データサイエンス人材育成の需要が高まる中で、注目度が上昇
- **合格率は6~7割**

G検定 累計受験者: 27,573名 累計合格者: 18,721名

G検定 累計受験者: 40,125名 累計合格者: 27,377名



2020年3回目は
受験者数 7,250名
合格者数 4,318名

<https://www.jdla.org/certificate/>

本講座の教材、画像、音声等の無断使用を固く禁じます

G検定の出題傾向 (概略)



- **人工知能(AI)、機械学習、ディープラーニング**の知識を
広範囲で満遍なく出題
- 出題範囲を把握し、偏りなく知識を吸収することが重要
- 常に技術が進化する分野なので、試験内容も毎回僅か変動する可能性がある

《全体的な傾向》

- 多く出る問題：
専門用語の定義、具体的な技術や分析手法の内容、人物名（技術の提案者）
- 少し出る可能性がある問題：
簡単な計算問題（微積分、統計処理、精度スコアの計算など）
- 意外と難しい問題：
AIに関する法律や規制、時事問題、最新技術の知見など

本講座の教材、画像、音声等の無断使用を固く禁じます

■ 人工知能（AI）

人工知能の定義や歴史、議論されている問題などに出てくる重要なワード

■ 機械学習

機械学習の基本概念、具体的な手法の名称、特徴、用途、精度評価の指標

■ ディープラーニングの概要と手法

ニューラルネットワークの仕組み、学習法と精度向上のための工夫、計算リソース、ディープラーニングの様々なテクニックと応用

■ ディープラーニングの研究分野と発展可能性

画像認識、自然言語処理、音声処理、ロボティクス、マルチモーダルなど、ディープラーニングの最新動向

■ ディープラーニングの応用に伴う社会的課題

現在議論されている法律や倫理についての広い知識が問われる

本講座の教材、画像、音声等の無断使用を固く禁じます

G検定試験を受験するメリット

【どんな方が受講するのか】

● 職場・学業に必要

- 人工知能、データ分析、機械学習、ディープラーニングのプロジェクトに携わる方、事業を開発する方

● 教養・スキルアップ

- AI・データサイエンスに興味を持ち、今後必要なスキルとして認識
- キャリアアップのために、自分が持つ知識を客観的な指標で証明したい方

【G検定を受験し合格すると嬉しいこと】

- 業務に役に立つ知識を幅広く、体系的学べる
- AIをビジネスに活用するための知識を有することを、肩書きで証明できる
 - 合格者に付与されるJDLAロゴマークを名刺や資料に掲載できる
- 受験に向けて講座、書籍、ウェブを活用して学習・調査していく中で、AIに関する豊富な情報や知識を集中的に整理できる

本講座の教材、画像、音声等の無断使用を固く禁じます

G検定出題分野の詳細解説

Part1：人工知能の定義・重要概念・技術の動向

■ 人工知能の定義・重要概念

- AIの定義
- AIの重要な概念：機械学習、ルールベース、ディープラーニング
- AIの4つの技術レベル

■ 人工知能の技術の動向

- AI技術の発展の歴史、関与した研究者たち
- 探索・推論、知識表現、機械学習、ディープラーニング などそれぞれのAI分野に関する技術的な内容

【一言アドバイス】

- 固有名詞など純粹な「知識」が出題されやすい**暗記中心の分野**
- **しっかり覚えておけば、試験で比較的点数を取りやすい**
- 記憶を助けるためにチートシートや単語帳を自作することがおすすめ

人工知能の研究開発の中で、今まで直面してきた問題

- トイ・プロブレム
- チューリングテスト
- シンボル・グラウンディング問題
- フレーム問題
- 身体性
- 弱いAIと強いAI
- シンギュラリティ

キーワード

【一言アドバイス】

- AIと人間の知能はどう異なるのかに関する概念が多い
- 類似している概念もあり、ひっかけ問題が多い分野
- 各「問題」の端的な定義を繰り返し復唱し、しっかり覚えましょう

本講座の教材、画像、音声等の無断使用を固く禁じます

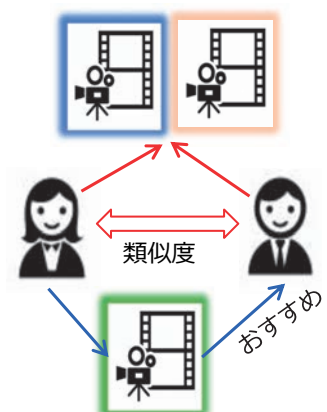
Part2 : 機械学習

- 「機械学習モデルを構築する流れ」「機械学習で目指すこと」などの機械学習の本質に関わる内容
- 機械学習の各分野の特徴（教師あり学習、教師なし学習、強化学習）
- 代表的な分析手法の特徴、仕組み、使い道
- 学習済みモデルの精度評価、精度指標
- 精度検証のためのデータの分割法
- ハイパーパラメータチューニング

購買履歴の分析



レコメンド機能



本講座の教材、画像、音声等の無断使用を固く禁じます

G検定試験で問われやすい代表的な機械学習手法

【教師あり学習】

- 線形回帰・ロジスティック回帰
- 決定木、ランダムフォレスト、勾配ブースティング
- K近傍法
- ナイーブベイズ法
- サポートベクターマシン (SVM)
- ニューラルネットワーク

【教師なし学習】

- K-平均法 (K-means)
- 主成分分析

本講座の教材、画像、音声等の無断使用を固く禁じます

Part2・機械学習：一言アドバイス

- 機械学習には様々な手法があり、内部の仕組みなどかなり具体的な内容が問われる問題が多い
- 初学者、文系出身者はこの分野に多めの学習時間を当てましょう

- 手法の設計コンセプトや特徴が問われやすい
- 各種の評価指標の特徴や使い分けが出題されやすい
- 簡単な計算を行い、モデルの性能を算出する問題も出題される可能性がある

改めて機械学習の目的について考えつつ、各手法について以下に注目

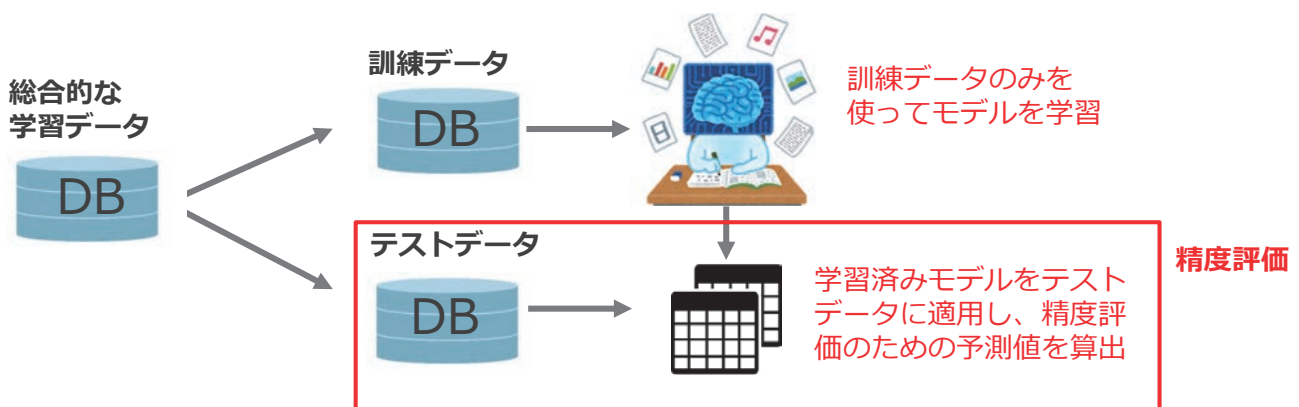
- 本質的な特徴、メリットとデメリット
- 活用事例を通じて、何故このように活用されているのか

各手法を自分の言葉で、具体的に説明できるかを自己チェック！

本講座の教材、画像、音声等の無断使用を固く禁じます

（復習）機械学習の目的は、学習データから汎用的なパターンを見出すように学習を行い、未知データに対しても精度よく予測できるようになること

- 未知データに対して高精度に予測する能力は**汎化性能**という
- 学習済みモデルを本番環境で使う前に、その汎化性能を証明することが重要
- 精度を検証する際にまず学習データを2部分に分ける
 - 訓練データ**（学習データと呼ぶこともある）：モデルを学習するため
 - テストデータ**：汎化性能を評価するため



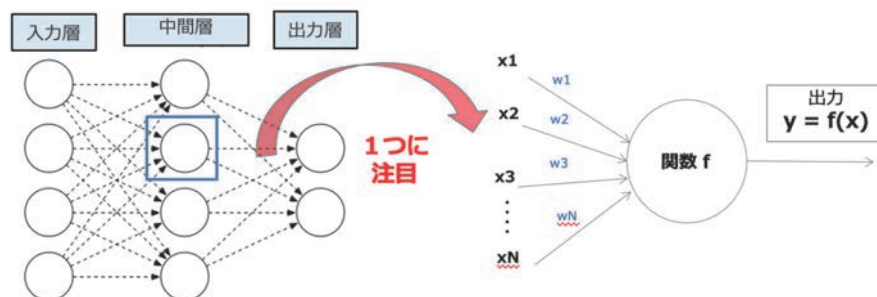
本講座の教材、画像、音声等の無断使用を固く禁じます

Part3 : ディープラーニング

■ ディープラーニングの概要

ニューラルネットワークの仕組みを理解するための基礎知識が問われる

- ニューラルネットワークの仕組みと学習法
- 既存のニューラルネットワークにおける難点と実用化に向けたアプローチ
- 計算リソース（CPU、GPU）などコンピュータに関する知識



■ ディープラーニングの代表的なテクニック

- 精度向上のための工夫
- 画像認識の王道手法であるCNNモデル
- ディープラーニングの具体的なモデルとその特徴
- ディープラーニングの応用分野（RNN、深層強化学習、深層生成モデル）

本講座の教材、画像、音声等の無断使用を固く禁じます

Part3・ディープラーニング：一言アドバイス

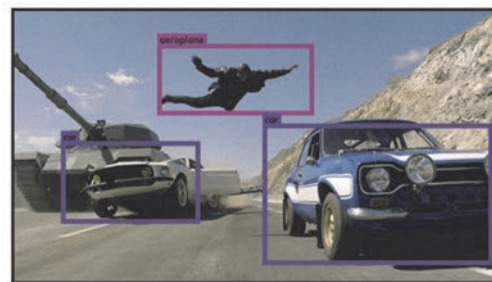
- Part3の内容ボリュームと出題数が大きい
→ 演習の時間配分をPart3に多めに割くと良いでしょう
- 理論を1つずつ深く掘り下げる必然性はなく、G検定にとって重要な項目の本質的な理解および体系的に技術を概観することに注力してください
- Part3の後半に位置する、RNN、強化学習、生成モデルの分野に関しては、具体的なモデルの名称や内部仕組みが問われることもあり、比較的難易度が高く点数も取りにくい
- 簡単目な微積分の問題が出題されたこともある（ニューラルネットの学習法に関係するため）
- ディープラーニングは魅力的な発展を遂げている分野なので、興味を持った内容について積極的にウェブで調べると理解を深めやすくなる

本講座の教材、画像、音声等の無断使用を固く禁じます

Part4：ディープラーニングの研究分野

- ディープラーニングが応用されている様々な研究分野について、Part3よりも浅く広く問われる

- 画像認識、一般物体検出
- 音声生成・音声処理
- 自然言語処理
- ロボティックス（強化学習）
- マルチモーダル



【一言アドバイス】

- Part3ほど深掘りせず、基本的に大枠的な内容と各技術分野の最新動向を捉えると良い
- ただ、たまには細かいことが問われることもあり、特に自然言語処理と強化学習に関しては、アルゴリズムの中身が問われやすい
- 書籍などでは殆ど説明されない「マルチモーダル」も近年出題されやすい

本講座の教材、画像、音声等の無断使用を固く禁じます

自然言語処理の分野で取り上げられる用語（一例）

形態素解析、Bag-of-Words (BoW) 、TF-IDF

One-Hotベクトル表現

分散表現、単語埋め込みモデル

Word2Vec、スキップグラム、CBOW

fastText、ELMO、BERT

トピックモデル

系列変換モデル (seq2seq)

Attention

GNMT (Google Neural Machine Translation)

ニューラル画像脚注付け

本講座の教材、画像、音声等の無断使用を固く禁じます

Part5 ディープラーニングの応用に伴う社会的課題

G検定では技術的な知識だけではなく、

AI関連分野の法律・規則についても出題される

- ディープラーニングの社会・産業への応用
- 最新技術に関する規制（例：自動運転車、ドローン）
- データサイエンスに深く関係する法律（特許、著作権、個人情報など）

試験対策で軽視されやすく、**合否の決め手になることも！**

AI分野の仕事に携わる全員が知っておく知識が多い

【一言アドバイス】

- 近頃、出題数が増えており、難易度も上がっている
- この分野では、いかに満遍なく演習問題をこなし、法律の細かい内容を覚えられるかによって、点数にその差が現れる
- 身近な話題が多いので、日頃から最新動向に意識を向けたり、想像力を持って学習すると、知識を覚えやすくなる

本講座の教材、画像、音声等の無断使用を固く禁じます

Part5で出題されやすい用語例

<2020年3月の試験後、Twitterにおける最頻出用語>

- 知的財産法
- 特許発明の要件
- 著作権
- 不正競争防止法
- 限定提供データ
- 個人情報保護法
- 匿名加工情報

キーワード	登場回数
個人情報（保護）	36
法律・倫理	28
営業秘密	25
XAI	12
音声認識	6
時事問題	3
強化学習	3
統計学	2

試験日（2020年11月7日）から1か月間にわたるtwitterの投稿を「G検定」で検索をかけてデータを収集

本講座の教材、画像、音声等の無断使用を固く禁じます

試験問題のイメージ



問題：第2次AIブームの特徴として、適切なものを1つ選べ

1. 推論と探索を可能とするAI
2. 大量な知識を蓄え、専門家のように返答できるAI
3. AIによって、特徴表現を自動的に抽出可能になった
4. AIが迷路やチェスなど明確なルールのある問題に対応可能になった

問題：オントロジーに関して正しくないものを1つ選べ

1. 知識の表現や共有の仕方に関する学問
2. AIにおいて概念を体系的に記述するための仕様
3. 言葉は哲学用語の「意味ネットワーク」から発祥
4. テキストマイニングに応用されることがある

問題：過学習の原因の候補として、適切ではないものを1つ選べ

1. モデルの関数が高次元で、複雑な形をしている
2. 学習データの量やバリエーションが多すぎる
3. 特徴量の数が非常に多い
4. お互いに相関が強い特徴量が一定数ある

本講座の教材、画像、音声等の無断使用を固く禁じます

問題：（ア）に当てはまるものとして、最も適切なものを1つ選べ

再帰型ニューラルネットワーク（RNN）は（ア）ことが可能であることが特徴の1つである。従来のRNNが抱える困難の1つとして（ウ）が挙げられる。現在RNNは（イ）などに活用されている。

1. 重要度の高い情報のみ残し、情報を効率よく圧縮する
2. 過去の入力を現在の出力に反映させる
3. 教師データを含まない学習データでも、分類に使用できる
4. 試行錯誤を通じて、報酬が最大となるような行動を見出す

問題：自然言語処理タスクの1つである形態素解析に関する記述として、最も適切なものを1つ選べ

1. 文の中の単語の間の係り受け関係を解析する
2. 各単語をそれに対応する1つの数値ベクトルに変換する
3. 文を意味を持つ最小単位となる単語に分割し、品詞を推定する
4. 主語、目的語など、単語間の関係性を解析する

本講座の教材、画像、音声等の無断使用を固く禁じます

短期間で効率的にG検定試験に合格

するための効果的な学習法

出題パターンに慣れる + 練習問題をたくさんこなす

- 自宅でオンライン試験中に、ウェブ検索やテキスト参照はOK
- 一方で、**出題数が多い**ため、いちいち調べていたら時間内に全回答することは難しい

では、何が必要なのか？

- **「数多くの練習問題をこなす」「身体で慣れる」**ことが必要
- 受験日までに十分に習熟度を達成し、知識が深く根付いていれば、問題文を見た瞬間考え込まずに**「自分の名前を答える」**と同じくらいすらすらと回答できるはず

**インプット（情報を読む）だけでは不十分で、
アウトプット（練習問題の数をこなす）がとても重要！**

本講座の教材、画像、音声等の無断使用を固く禁じます

おすすめ学習法#2

知識の覚え方を極める

暗記勝負のG検定 …

記号のような丸暗記だと苦痛なので、以下がおすすめ

- テキストを読む/講義を視聴する時 **「理解した！」**と実感できるまで**同じ場所を繰り返して**ください
- 仮想的な質問者に **「自分の言葉で相手に説明できるか」**を試す
- 「覚える」ために**同じ問題を繰り返し解くのも意味がある**
- **「どうしても覚えきれない所だけをまとめたノート」「フラッシュカード」**を作る（本番直前に**「これだけ見れば良い！」**モノ）

本講座の教材、画像、音声等の無断使用を固く禁じます

アガルートアカデミーでは、
G検定試験対策講座を提供しています

<https://www.agaroot.jp/datascience/gkentei/>

- **基本概念から一步一步親切丁寧に**解説していき、初めてデータサイエンスを学ぶ方でも安心できます
- **基礎からG検定合格にリーチするレベルまで**学習したい方に最適な講座



直前対策版のライト版もご用意してあります！

本講座の教材、画像、音声等の無断使用を固く禁じます

最後に ...



- G検定には難易度の高い出題もみられるが、必ずしも受験の際に全てを知っていることが期待されているわけではない
- 「**受験を機に、解けなかった問題については自主的に調べて知識を広げてください**」というメッセージのように受け止められる
(個人の推測)

合格はゴールではなく、ジェネラリストとしてのスタート

継続的に勉強を続ける姿勢、それこそがジェネラリストに求められること！

本講座の教材、画像、音声等の無断使用を固く禁じます