似た分子から物性を予測する

分子設計 第 3 回 ｜ 教科書 第 3 章

第 3 話　似た分子から見積もる

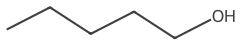
「似ている」をどう定義すると、予測がどう変わるか

これまでの経過

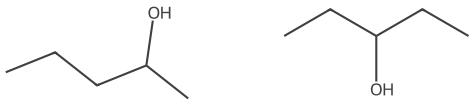
第 2 回　分子のどの特徴を、表の列にするか

C5 の 5 分子 → (n, s, b) の行は 3 種類

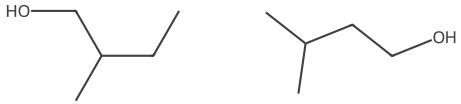
$(5, 0, 0)$



$(5, 1, 0)$



$(5, 0, 1)$



今日できるようになったこと

- ○ 骨格式・縮合構造式から (n, s, b) を数えられる
  OH の炭素の級と分岐点を数える。OH との結合は数えない
- ○ 同じ (n, s, b) の分子を見つけ、
  表せない違い（OH の位置・分岐の位置）を言える
- ○「区別できる」と「予測が合う」を分けて言える

まだ分からないこと

- ？ 記述子の近い分子の沸点から、
  候補分子の沸点を予測できるか → 第 3 回
- ？ 同じ行の分子の沸点は、どれだけ違うか → 第 5 回

依頼沸点 115〜140 ℃ ｜ w_{25} 4.0 以上 ｜ C5〜C6 の一級・二級

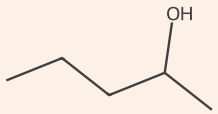
前回の確認 2-メチル-1-ブタノール（$HOCH_2CH(CH_3)CH_2CH_3$）の s は　（0／1）

s：OH の付いた炭素が二級なら 1、一級なら 0

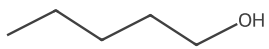
問い 1　「似ている」をどう定義すると、予測がどう変わるか



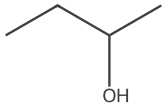
工程担当者　「構造を変えると、沸点も範囲から外れないか心配。
資料が届く前に、2-ペンタノールの沸点を見積もれる？」



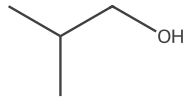
2-ペンタノール
（予測したい）
$(5, 1, 0)$
沸点 ？



1-ペンタノール
$(5, 0, 0)$
沸点 138 ℃



2-ブタノール
$(4, 1, 0)$
沸点 100 ℃



2-メチル-
1-プロパノール
$(4, 0, 1)$
沸点 108 ℃

やってみる　どの分子の値を使いたいか、1 つ選ぶ 着目点：炭素数、OH の炭素の級、分岐

今日の流れ

教科書 §3.1〜3.6

依頼：1-ペンタノールを改良し、候補と次の実験を提案する（第 1〜6 回）

データを整える	構造を表す	予測する		確かめる	候補と次の実験
第 1 回 ✓ データ	第 2 回 ✓ 記述子	**第 3 回 似た分子**	第 4 回 予測式	第 5 回 検証	第 6 回 提案

▲ 今日

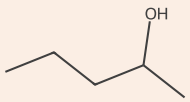
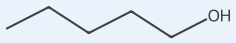
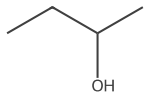
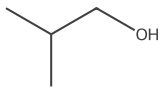
1. **問い 1 ①**　「似ている」を、どう数にするか → 例題・基本演習 A
2. **問い 1 ②**　重みを変えると、予測はどう変わるか → 例題・基本演習 B
3. **問い 1 ②**　どの予測を採用するか → 応用演習

そのあと：問い 1 の答え → 実際のケモインフォでは → 自分の実験では

似た分子の値を使う：最近傍による予測

教科書 §3.1

問い 1 ①　「似ている」を、どう数にするか

参照分子		参照分子の炭素数 n'	沸点 / ℃	距離
	**2-ペンタノール**（予測したい）	$n = 5$	?	
	1-ペンタノール	5	**138**	$\lvert 5-5 \rvert = 0$ **最近傍**
	2-ブタノール	4	**100**	$\lvert 5-4 \rvert = 1$
	2-メチル-1-プロパノール	4	**108**	$\lvert 5-4 \rvert = 1$

→ 2-ペンタノールの予測

138 ℃

最近傍の値を使う

- **距離**：記述子の違いから決める、似ていなさの数。小さいほど似ている
- **最近傍**：距離が最も小さい参照分子。その値を予測値にする

構造の違いを、どう距離にするか

教科書 §3.2・§3.3

問い 1 ①　「似ている」を、どう数にするか

$$d = |n - n'| + 2|s - s'| + |b - b'|$$

重み：違い 1 つを、距離いくつと数えるか（s の 2 は仮の値。② で変える）

違い 1 つ	例	距離への寄与	
n が 1 違う	炭素が 1 つ多い・少ない		1
b が 1 違う	分岐点が 1 つ多い・少ない		1
s が 1 違う	一級と二級（OH の炭素の級）		2 重み 2

この授業の値の幅　n 2〜6（幅 4）　s 0 か 1　b 0 か 1

例題：2-プロパノールの最近傍

問い 1 ①　距離をどう計算するか

問題　2-プロパノール $(3, 1, 0)$ の沸点を、s の重み 2 の距離で予測してください。参照分子：
1-ペンタノール $(5, 0, 0)$・2-ブタノール $(4, 1, 0)$・2-メチル-1-プロパノール $(4, 0, 1)$

解き方

① 1-ペンタノール　$|3-5| + 2|1-0| + |0-0|$

$$= 2 + 2 + 0 = 4$$

② 2-ブタノール　$|3-4| + 2|1-1| + |0-0| = 1$

③ 2-メチル-1-プロパノール　$|3-4| + 2|1-0| + |0-1| = 1 + 2 + 1 = 4$

答え　最近傍は 2-ブタノール（距離 1）。予測 100 ℃

基本演習 A：s の重み 2 で予測する

問い 1 ①　距離をどう計算するか

2-ペンタノール $(5, 1, 0)$　$d = |n - n'| + 2|s - s'| + |b - b'|$

手順 1　1-ペンタノールまでの d を手で：違いを n・s・b の順に → s に 2 → 足す

手順 2　最近傍を予想してから、アプリで残り 2 つを確かめる

手順 3　最近傍の値から、予測値を読む

問　炭素数だけの予測（138 ℃）から変わったのは、何の違いを重く見たからか

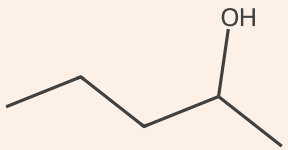
2-ペンタノール $(5, 1, 0)$　n　2 × s　b の違い　$d = |n - n'| + 2\,|s - s'| + |b - b'|$

1-ペンタノール $(5, 0, 0)$	**138 ℃**		$d = ?$
2-ブタノール $(4, 1, 0)$	**100 ℃**		$d = ?$
2-メチル-1-プロパノール $(4, 0, 1)$	**108 ℃**		$d = ?$

予測 **？**

基本演習 A：残り 2 つをアプリで

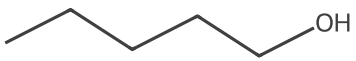
最近傍を 1 つ押して予想 →「距離を計算」



2-ペンタノール
（予測したい）

$(5, 1, 0)$

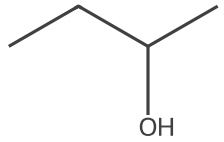
沸点 ？



1-ペンタノール

$(5, 0, 0)$

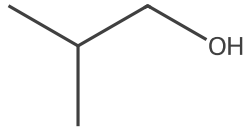
沸点 138 ℃



2-ブタノール

$(4, 1, 0)$

沸点 100 ℃



2-メチル-
1-プロパノール

$(4, 0, 1)$

沸点 108 ℃

$d = |n - n'| + 2|s - s'| + |b - b'|$ で 3 つの参照分子までの距離を求め、最近傍の沸点を予測値にする

基本演習 A の答え（1）

手順 1　1-ペンタノールまでの距離

① 違い　n：0、s：1、b：0

② 足す　$0 + 2 \times 1 + 0 = 2$

答え　$d = 2$

手順 2・3　最近傍と予測

考え方　距離：1-ペンタノール 2、
2-ブタノール 1、
2-メチル-1-プロパノール 4

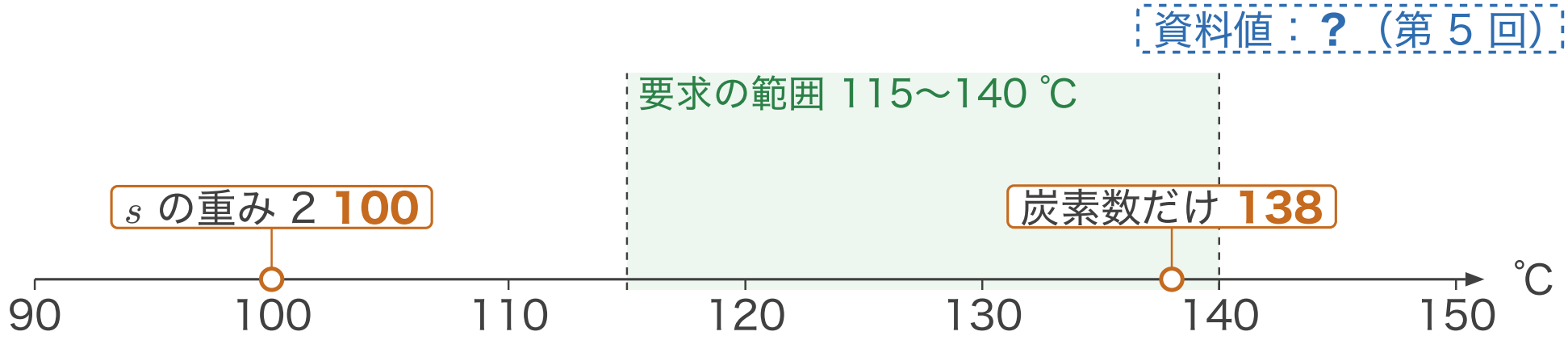
答え　**最近傍は 2-ブタノール。予測 100 ℃**

基本演習 A の答え（2）

問　炭素数だけの予測（138 ℃）から変わった理由

考え方　炭素数だけでは 1-ペンタノールが距離 0。
s の違いを距離に入れ、重み 2 で炭素 1 個の違いより重く数えた

答え　OH の炭素の級（s）の違いを、重く見るようになったから

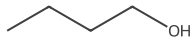


要点：「似ている」の決め方を変えると、最近傍も予測も変わる（138 → 100 ℃）

同じ距離になったら、どうするか

教科書 §3.4

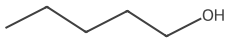
問い 1 ② 同じ距離になったら、どうするか



1-ブタノール

$(4, 0, 0)$

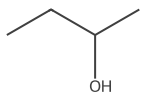
どれに似ているか



1-ペンタノール

$(5, 0, 0)$

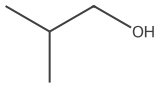
炭素が 1 つ多い



2-ブタノール

$(4, 1, 0)$

OH の炭素が二級



2-メチル-1-プロパノール

$(4, 0, 1)$

分岐点が 1 つ

1-ブタノールからの距離	n の違い	s の違い	b の違い	最近傍
s の重み 2	1	2	1	同じ距離が 2 つ → ?
s の重み 1	1	1	1	同じ距離が 3 つ

例題：同じ距離の最近傍が 2 つあるとき

問い 1 ②　同じ距離になったら、どうするか

問題　1-ブタノール $(4, 0, 0)$ の沸点を、(1) 1-ペンタノール $(5, 0, 0)$・(2) 2-ブタノール $(4, 1, 0)$・(3) 2-メチル-1-プロパノール $(4, 0, 1)$ から、s の重み 2 で予測してください

解き方

① 距離　(1) 1　(2) $0 + 2 \times 1 + 0 = 2$　(3) 1

② 最近傍　(1) 1-ペンタノールと
(3) 2-メチル-1-プロパノール

③ 平均　$(138 + 108)/2 = 123$

答え　予測 123 ℃（同じ距離の 2 つを平均）

要点：同じ距離の最近傍は、すべての値を平均する

基本演習 B：s の重み 1 で予測する

問い 1 ②　重みを変えると、予測はどう変わるか

2-ペンタノール $(5, 1, 0)$　$d = |n - n'| + |s - s'| + |b - b'|$

手順 1　3 つの参照分子までの距離を手で求める

手順 2　最小が 2 つあれば、平均して予測値を出す

手順 3　次のスライドのアプリで確かめる

問 1　重み 2 の予測（100 ℃）と比べて、予測は　（変わる／変わらない）

問 2　3 つの予測のどれが正しいかを、今の情報で　（決められる／決められない）

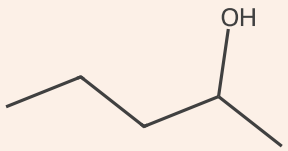
2-ペンタノール $(5, 1, 0)$　n　1 × s　b の違い　$d = |n - n'| + 1\ |s - s'| + |b - b'|$

1-ペンタノール $(5, 0, 0)$	**138 ℃**		$d = ?$
2-ブタノール $(4, 1, 0)$	**100 ℃**		$d = ?$
2-メチル-1-プロパノール $(4, 0, 1)$	**108 ℃**		$d = ?$

予測 **？**

基本演習 B：アプリで確かめる

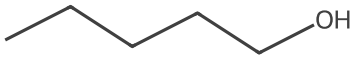
手で解いたら、最近傍を 1 つ押して「距離を計算」。同じ距離のときの予測値を確かめる



2-ペンタノール
（予測したい）

$(5, 1, 0)$

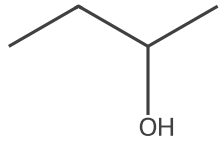
沸点 ？



1-ペンタノール

$(5, 0, 0)$

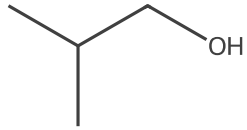
沸点 138 ℃



2-ブタノール

$(4, 1, 0)$

沸点 100 ℃



2-メチル-
1-プロパノール

$(4, 0, 1)$

沸点 108 ℃

s の重みを 1 にして 3 つの距離を求める。同じ距離の最近傍は、すべての沸点を平均する

基本演習 B の答え（1）

手順 1・2　s の重み 1 の最近傍と予測

① 距離　1-ペンタノール 1、2-ブタノール 1、2-メチル-1-プロパノール 3

② 平均　同じ距離の 2 つ：$(138 + 100)/2 = 119$

答え　予測 119 ℃

2-ペンタノール $(5, 1, 0)$　　n 1 × s b の違い　　$d = |n - n'| + 1\ |s - s'| + |b - b'|$

1-ペンタノール $(5, 0, 0)$	**138 ℃**	s 1	$d = 1$ **最近傍**
2-ブタノール $(4, 1, 0)$	**100 ℃**	n 1	$d = 1$ **最近傍**
2-メチル-1-プロパノール $(4, 0, 1)$	**108 ℃**	n 1　s 1　b 1	$d = 3$

同じ距離の最近傍は平均 → $(138 + 100)/2$ ＝ 予測 **119 ℃**

基本演習 B の答え（2）

問 1　重み 2 の予測と比べて

考え方　100 ℃ → 119 ℃。重みを変えただけで、最近傍と予測が変わる

答え　変わる

問 2　3 つの予測のどれが正しいか、今の情報で決められるか

考え方　2-ペンタノールの資料値をまだ見ていない。3 つとも、決め方から出た予測

答え　決められない



要点：同じ距離なら平均する。予測の違いは、決め方の違いから出ている

理解確認（個人・3 分）

$d = |n - n'| + 2|s - s'| + |b - b'|$、予測したい分子は $(4, 1, 0)$

- 参照分子 U：$(3, 1, 0)$、沸点 90 ℃　　参照分子 V：$(4, 0, 0)$、沸点 110 ℃（架空）

1. U・V までの距離を求め、最近傍の値を使った予測を示す
2. 「最近傍が選べたので、この予測値は正しい」と言えるか　（言える／言えない）
   理由を 1 文で

予測したい分子 $(4, 1, 0)$　　n　2 × s　b の違い　　$d = |n - n'| + 2\,|s - s'| + |b - b'|$

参照分子 U $(3, 1, 0)$	90 ℃		$d = ?$
参照分子 V $(4, 0, 0)$	110 ℃		$d = ?$

予測 ？

理解確認の答え

問 1　U・V までの距離と予測

考え方　$\mathrm{U}: 1 + 2 \times 0 + 0 = 1$、$\mathrm{V}: 0 + 2 \times 1 + 0 = 2$

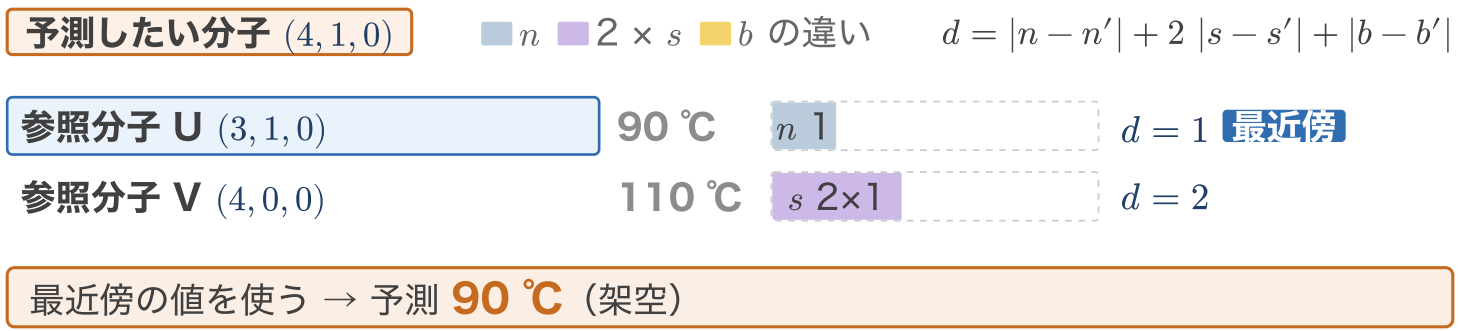
答え　最近傍は U。予測 90 ℃（架空の値）

問 2　「最近傍が選べたので、この予測値は正しい」と言えるか

考え方　距離は構造の近さの決め方で、沸点の一致は保証しない

よくある誤り　「計算ミスかもしれない」だけでは理由にならない（計算が正しくても言えない）

答え　言えない



応用演習 1：重みを少しずつ変える

問い 1 ②　重みを少し変えると、どうなるか

s の重みを α とする。距離は下の横棒のとおり

手順 1　　$\alpha = 0.9$ で、最近傍と予測値を出す

手順 2　　$\alpha = 1.1$ で、最近傍と予測値を出す

問　　重みを少しずつ変えると、予測は　（なめらかに変わる／急に変わる）

2-ペンタノール $(5, 1, 0)$　　n　$\alpha \times s$　b の違い　　$d = |n - n'| + \alpha|s - s'| + |b - b'|$

1-ペンタノール $(5, 0, 0)$	**138 ℃**	$\leftrightarrow \alpha$	$d = \alpha$
2-ブタノール $(4, 1, 0)$	**100 ℃**	n 1	$d = 1$
2-メチル-1-プロパノール $(4, 0, 1)$	**108 ℃**	n 1　$\leftrightarrow \alpha$　b 1	$d = 2 + \alpha$

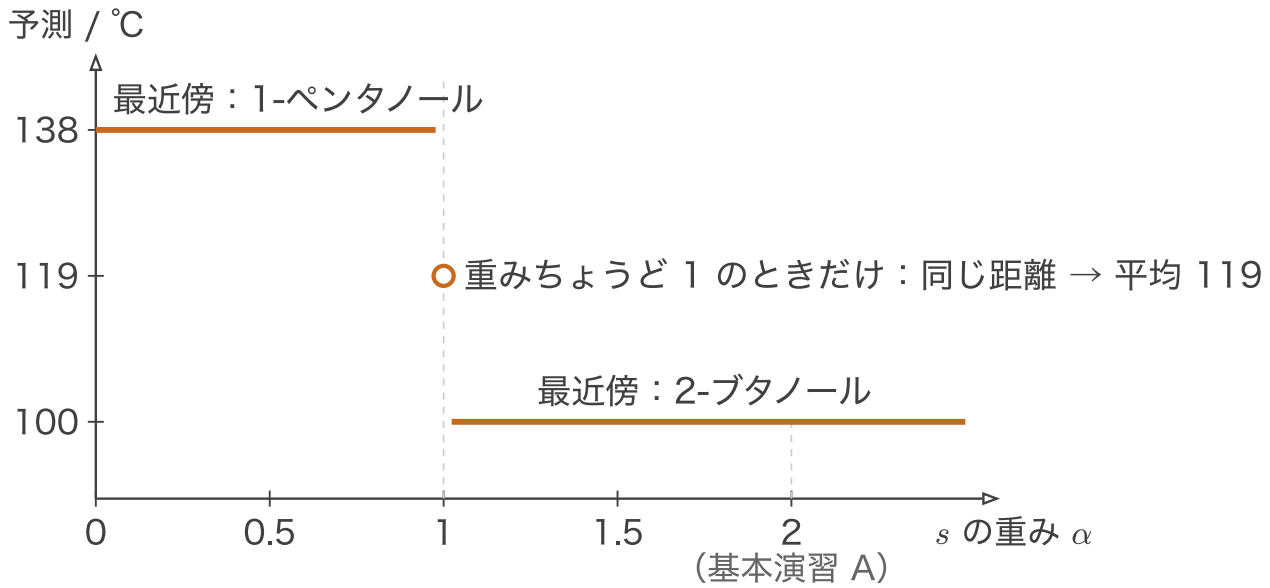
$\alpha < 1$ なら？　$\alpha = 1$ なら？　$\alpha > 1$ なら？

応用演習 1 の答え

問　重みを少しずつ変えると

考え方　n・b の重みは 1 のまま。距離は 1-ペンタノール α、2-ブタノール 1。
α が 1 より小さいと 1-ペンタノール、大きいと 2-ブタノールが最近傍

答え　急に変わる（0.9 で 138 ℃、1.1 で 100 ℃。119 ℃ は重み 1 のときだけ）



（基本演習 A）

応用演習 2：依頼人に答える

問い 1 ②　どの予測を採用するか

決め方	予測	何の違いを見たか	範囲に入るか
炭素数だけ	**138 ℃**		
s の重み 2	**100 ℃**		
s の重み 1	**119 ℃**		

手順 1　表の空いている 2 列を埋める

手順 2　答え：〔どの予測〕を採用する。〔何の違い〕を見た。未確認は〔何か〕

範囲は 115〜140 ℃。正しさではなく、何の違いを見るかで選ぶ。「決めない」でもよい

応用演習 2 の答え（1）

表　何の違いを見たかと、範囲に入るか

答え　決め方で、範囲に入るかまで変わる

炭素数だけ

似ているとした分子

1-ペンタノール

重く見た： 炭素数の違い

予測 **138 ℃**

115〜140 ℃ に：**入る**

s の重み 2

似ているとした分子

2-ブタノール

重く見た： OH の炭素の級の違い

予測 **100 ℃**

115〜140 ℃ に：**下回る**

s の重み 1

似ているとした分子

1-ペンタノール
2-ブタノール

重く見た： 炭素数と級を同じに

予測 **119 ℃**（平均）

115〜140 ℃ に：**入る**

応用演習 2 の答え（2）

答え　いまの時点で、どの予測を採用するか

成立する例　138 ℃：炭素数の違いが沸点に最も影響すると見た

成立する例　100 ℃：級の違いを、炭素 1 個の違いより重く見た

成立する例　119 ℃：炭素数の違いと級の違いを、同じ重みで見た

成立する例　決めない：3 つとも決め方から出た予測。資料値で確かめてから選ぶ

成立しない例　「真ん中の予測だから正しい」「この重みが最良の重み」：
確かめていないことを結論にしている

答え　理由と、未確認のこと（2-ペンタノールの資料値）が書けていれば成立する

要点：予測は決め方しだいで変わる。どれがよいかは、確かめるまで言わない

問い 1 の答え　記述子と重みの決め方で、最近傍も予測値も変わる

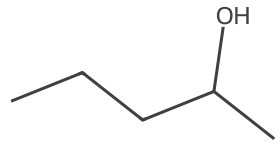
問い 1：「似ている」をどう定義すると、予測がどう変わるか



- 「似ている」＝記述子の違いから決めた距離。最近傍の値を使い、同じ距離なら平均
- どれを採用しても、範囲に入るかは資料値で確かめるまで分からない（第 5 回）

実際のケモインフォでは：似た分子を探し、その値を使う

調べたい分子



1 0 1 0 0 1

フィンガープリント
（第 2 回）

→

データベースの分子と比べる

類似度（Tanimoto 係数）
＝ 共通して 1 の数
　÷ どちらかが 1 の数

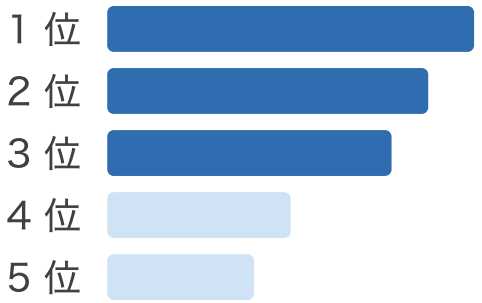
調べたい 1 0 1 0 0 1
比べる 1 0 0 0 1 1

共通して 1：2　どちらかが 1：4
$\rightarrow 2/4 = 0.5$

0〜1。1 に近いほど似ている
（距離とは向きが逆）

→

似ている順に並べる



上位 k 個の値の平均 → 予測
（k 近傍法。k は自分で決める数）

この授業：3 分子との距離、最近傍（同じ距離なら平均）

実際の研究：多数の分子と比べ、上位 k 個を使う

「似ている」を数で決めると、予測にも検索にも使える

自分の実験では、何を似ているとするか

合成：新しい基質の収率を、既知の基質から見積もる

似ている　骨格と、置換基の位置
違う　置換基の種類、反応の温度・時間

材料：新しい組成の性質を、近い組成から見積もる

似ている　主な元素と、組成比の近さ
違う　焼成温度、試料の形、測定の温度

- 既知の結果を参照するときは、何を似ているとしたか、何が違うかを書いておく
- 自分の研究（まだなら学生実験）で参照している既知の例を 1 つ挙げ、
  似ている点と違う点を考えておく（宿題でよい）

まとめ：今日できるようになったこと

今日できるようになったこと（到達目標 ③）

1. 記述子の違いから距離を計算し、
   最近傍の値を使って予測できる
   同じ距離の最近傍は、すべての値を平均する
2. 重み（何の違いを重く見るか）を変えると、
   最近傍も予測も変わることを示せる
   2-ペンタノール：炭素数だけ 138 ℃、
   重み 2 で 100 ℃、重み 1 で 119 ℃
3. 重みを少し変えると、最近傍が入れ替わり、
   予測が急に変わることを説明できる。
   最近傍が選べても、予測が正しいとは言えない

まだ分からないこと

- 似た分子を選ぶ代わりに、
  構造と沸点の傾向を式にできるか
  → 第 4 回
- 3 つの予測のどれが資料値に近いか
  第 5 回に資料値と照らし合わせる

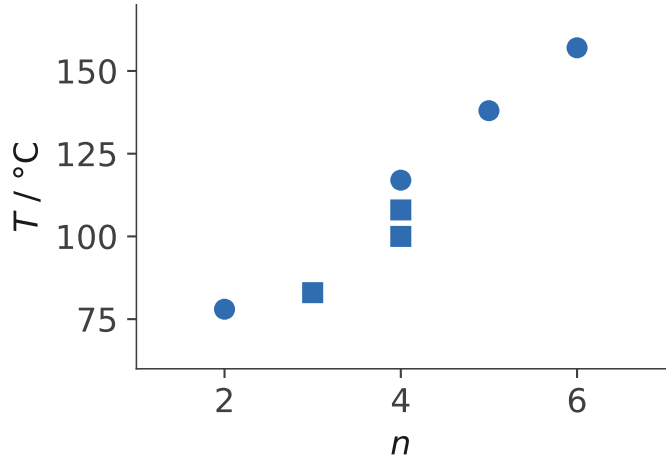
次回予告：第 4 回 データから予測式を作る

教科書 第 4 章

第 4 話　式にまとめ、予測を封印する



工程担当者　「使う分子しだいで答えが変わるのは困る。式で言える？」



T：沸点、n：炭素数。丸：直鎖一級 4 分子
四角：二級・分岐 3 分子（2-プロパノールは 83 ℃）

次回の問い：構造と沸点の傾向を、簡単な式にまとめられるか

今日の復習 小テスト第 3 回（manaba）　**教科書で読む** 第 3 章（§3.2 の発展は読むだけ）　**予習** §4.1〜4.2