

# Лабораторийн даалгавар 6

Даалгавар өгсөн: 2025 оны 5-р сарын 5, Даваа гараг.

**Хугацаа: 2025 оны 5-р сарын 19-ны Даваа гараг**

## Удиртгал

2012 онд шинээр сонгогдсон Ерөнхийлөгч Дональд Трамп уур амьсгалын өөрчлөлтийг Хятад улс зохиосон луйвар хэмээн итгэдэгээ илэрхийлж байсан. Энэ даалгаварт бид түүнийг буруу гэдгийг батлахыг оролдох болно. Бид регрессийн шинжилгээ ашиглан АНУ-ын янз бүрийн бүс нутгийн уур амьсгалыг загварчлан, дэлхийн дулаарлын нотолгоог олохыг зорьж байна.

- Эхлээд та температурын хувьд уур амьсгалын өөрчлөлтийг шинжилж, дүрслэх загваруудыг үүсгэнэ. Мөн мэдээллийг илүү ойлгомжтой болгохын тулд шуугиан багатай болгох арга замуудыг авч үзнэ.
- Дараа нь таны боловсруулсан загварууд түүхэн өгөгдөл ашиглан ирээдүйн температурыг хэр сайн таамаглаж чадаж байгааг шалгана.
- Сүүлд нь температурын ердийн бус, хэт хэлбэлзлийг (зөвхөн өсөлт бус) загварчлах аргыг судална.

Бидний өгсөн файлуудын нэрийг бүү өөрчил, өгөгдсөн туслах функцүүдийн аль нэгийг бүү өөрчил, функц/методын нэрийг бүү өөрчил, эсвэл өгөгдсөн docstrings-ийг бүү устга. Та data.csv файлыг ps5.py файлтай ижил хавтсанд хадгалах шаардлагатай. Мөн Хэв маягийн гарын авлагыг үзнэ үү, учир нь бид тодорхой зөрчлүүдийн хувьд оноо хасах болно. Цаашилбал, тодорхойгүй хувьсагчийн нэр болон тайлбаргүй кодын хувьд хасагдах болно. Үүнээс гадна, та rpylab баримт бичгийг шалгаж үзэж болно (энд байгаа numpy болон scipy хэсгүүдийг үзнэ үү), учир нь rpylab нь энэ даалгаврыг илүү хялбар болгох хэд хэдэн функцийг агуулдаг.

Та дараах зүйлсийг илгээх ёстой:

1. ps5.py дахь шийдлүүдийн код
2. ps5\_writeup.pdf дахь хэлэлцүүлгийн бичлэг

data.csv файлыг бүү илгээгээрэй, учир нь та үүнд ямар ч өөрчлөлт оруулахгүй.

## Бичвэр

Энэ даалгаврын хувьд та өөр өөр бодлогуудын үр дүнд үүссэн график дээр тулгуурласан тайлбар бичиг (write-up) ирүүлэх шаардлагатай.

- Та өөрийн бичсэн кодоор үүсгэсэн графикуудыг энэ тайлбар бичигт заавал оруулах ёстой.
- Хэрэв та графикууд тань буруу болсон гэж бодож байсан ч, өгөгдсөн график/кодын дагуу асуултад хариулна уу. Мөн та өөрийн бодлоор ямар үр дүн гарах ёстой байсан талаар тайлбар оруулж болно.

Энэ даалгавар нь өгөгдөл дэх хандлагыг судлах зорилготой учраас, бид тайлбар бичгийг таны дүн шинжилгээний чанар дээр тулгуурлан илүү нарийвчлалтайгаар үнэлэх болно.

## Climate Класс

Тодорхой бүс нутгийн уур амьсгалын өөрчлөлтийг загварчлахын тулд танд өгөгдөл хэрэгтэй болно. Энэ бодлогын багцын хүрээнд бид АНУ-ын Байгаль орчны мэдээллийн үндэсний төв (NCEI)-ээс авсан температурын өгөгдлийг ашиглана.

data.csv файлд 1961 оноос 2015 он хүртэлх хугацаанд АНУ-ын 21 хотын өдөр бүрийн хамгийн их ба хамгийн бага температурын хэмжилтийг агуулсан өгөгдөл байгаа.

Энэ файлыг нээж, түүхий өгөгдлийг (raw data) хараарай.

Түүхий өгөгдлийг боловсруулахын тулд ps5.py файлд бид танд туслах зориулалттай Climate гэсэн класс бэлдсэн. Та энэ классын объект үүсгэхдээ өгөгдлийн файлын нэрийг оруулж өгнө.

🔗 Энэ классыг сайтар уншиж танилцаад, docstring-уудыг ашиглан дараагийн бодлогуудын өгөгдлийг хэрхэн авахыг ойлгоорой.

## А хэсэг: Загваруудыг Үүсгэх

### Бодлого 1: Муруйг Тохируулах

generate\_models функцийг хэрэгжүүлнэ үү. Энэ функц нь (x,y) координатуудын багц болон зэргүүдийг (1=шугаман, 2=квадрат, 3=куб гэх мэт) авч, өгөгдлийн цэгүүдэд заасан зэргийн полиномуудыг тохируулдаг. Дараа нь хамгийн сайн тохирсон полином бүрийн коэффициентуудыг буцаана.

Зөвлөмж: `pylab.polyfit`-ийн баримт бичгийг үзнэ үү.

Функцийн оролтууд:

x болон y нь өгөгдлийн түүврийн x-координат болон y-координатуудтай харгалзах `pylab`-ийн нэг хэмжээст хоёр массив юм (Python жагсаалт биш). Жишээлбэл, хэрэв та N өгөгдлийн цэгтэй бол,

$x=[x_1, x_2, \dots, x_N]$  болон  $y=[y_1, y_2, \dots, y_N]$

энд  $x_i$  болон  $y_i$  нь  $i$  дэх өгөгдлийн цэгийн x болон y координатууд юм. Энэ бодлогот даалгаварт x-координат бүр нь бүхэл тоо бөгөөд түүврийн жилийн (жишээлбэл, 1997) харгалзана. Харгалзах y-координат бүр нь хөвөгч цэг бөгөөд тухайн жилийн температурын ажиглалтыг Цельсээр илэрхийлнэ (эдгээр ажиглалтыг хэрхэн яг тооцоолдогийг бид энэ бодлогот даалгаварын сүүлд тайлбарлах болно - та энэ функцийн хувьд санаа зовох хэрэггүй). Өгөгдлийн цэгүүдийн энэ хоёр массив бүхий илэрхийлэл нь бүх бодлогот даалгаварт ашиглагдана.

degс нь бүхэл тоонуудын лист бөгөөд бидний үүсгэх регрессийн загвар бүрийн зэргийг заана. Загвар бүрийн хувьд энэ функц нь өгөгдөл (x, y)-ыг тухайн зэргийн полином муруйд тохируулах ёстой.

Функцийн гаралт:

Функц нь загваруудын листийг буцаах ёстой. Загвар нь полиномын хувьд ашигласан коэффициентуудын 1-хэмжээст `pylab` массив юм. (Тэгэхээр таны эцсийн гаралт нь

pylab массивуудаас бүрдсэн лист юм.) Загварууд нь degs дахь харгалзах бүхэл тоонуудтай ижил дараалалд байх ёстой.

Жишээ:

```
print generate_models(pylab.array([1961, 1962, 1963]), pylab.array([-4.4, 5.5, -6.6]), [1, 2])
```

Дараахтай ойролцоо зүйл хэвлэх ёстой:

```
[array([-1.10000000e+00, 2.15270000e+03]), array([ 6.83828238e-14, -1.10000000e+00, 2.15270000e+03])]
```

Дээрх жишээ нь  $(x_i, y_i) = (1961, -4.4), (1962, -5.5),$  ба  $(1963, -6.6)$  өгөгдлийн түүврүүд дээр шугаман болон квадрат муруйг (зэрэг 1 ба 2) үүсгэдэг. Үр дүнд бий болсон загварууд нь degs-д заасан ижил дараалалд байна. Тоон алдаанаас болж яг таг тоо гарахгүй байсан ч зүгээр гэдгийг анхаарна уу. Үүнийг хэрэгжүүлсний дараа таны код test\_generate\_models нэгж тестийг давж өнгөрөх ёстой.

## Бодлого 2: $R^2$

Регрессийн хэд хэдэн загвар үүсгэсний дараа бид загваруудаа үнэлж, өгөгдөлд хэр зэрэг сайн тохирч байгааг олж мэдэн, хамгийн сайн загваруудыг сонгохыг хүсдэг. Загвар хэр сайн ажиллаж байгааг үнэлэх нэг арга бол загварын  $R^2$  утга буюу тодорхойлолтын коэффициент гэгддэг утгыг тооцоолох явдал юм.  $R^2$  нь түүврийн нийт хувилбар хэр зэрэг загвараар тайлбарлагдаж байгааг хэмжих хэмжүүр өгдөг. r\_squared функцийг хэрэгжүүлнэ үү.

Оролт:

у: бодит өгөгдлийн түүврийн у-координатуудыг агуулсан 1 хэмжээст pylab массив  
estimated: регрессийн загвараас үнэлсэн у-координатуудыг агуулсан 1 хэмжээст pylab массив

Гаралт:

Энэ функц нь тооцоолсон  $R^2$  утгыг буцаах ёстой.

Та  $R^2$ -ийг дараах байдлаар тооцоолж болно:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - e_i)^2}{\sum_{i=1}^n (y_i - \text{mean})^2}$$

Энд:

$e_i$  нь  $i$  дэх өгөгдлийн цэгийн хувьд үнэлэгдсэн (регрессээр таамагласан) у-утга  $y_i$  нь  $i$  дэх өгөгдлийн цэгийн хувьд бодит (анхны өгөгдлөөс) у-утга  $mean$  нь анхны өгөгдлийн түүврийн дундаж (өөрөөр хэлбэл,  $y_1, y_2, \dots, y_n$ )

Хэрэв та  $R^2$ -ийн талаар бүрэн ойлгоогүй хэвээр байвал, Wikipedia хуудас нь түүний хэрэглээ болон хэрхэн тооцоолох талаар сайн тайлбарласан байгаа. Зарим Python багцууд  $R^2$ -ийг шууд тооцоолох функцуудыг агуулдаг; та эдгээрийг ашиглаж болохгүй, тооцооллыг эхнээс нь хийх ёстой.

Зөвлөмж: `pylab` массивуудын тусламжтайгаар та давталт (`for-loop`) ашиглахгүйгээр олон үйлдлийг хялбархан гүйцэтгэх боломжтой. Жишээ нь:

Python

```
>>> a = pylab.array([1,2,3])
>>> b = pylab.array([1,0,1])
>>> a + b
array([2, 2, 4])
>>> a - 1
array([0, 1, 2])
>>> a * 2
array([2, 4, 6])
```

Таны код одоо `test_r_squared`-ийг давж өнгөрөх ёстой.

### Бодлого 3: Загваруудыг Дүрслэх

Бид үнэлгээ хийх тоон метрикийг хэрхэн олж авахыг сурсан. Өгөгдлийн түүврүүдээ тохирох муруйнуудтай хамт дүрслэх нь олж авсан загваруудын сайн эсэхийг тодорхойлоход тусална. Энэ бодлогод бид иж бүрэн үнэлгээ хийхдээ тоон метрик болон дүрслэлийг нэгтгэх болно.

`evaluate_models_on_training` функцийг хэрэгжүүлнэ үү. Энэ функц нь бидний загвар үүсгэхэд ашигласан ижил өгөгдөл дээр буюу сургалтын өгөгдөл дээр загваруудыг үнэлэхэд ашиглагдана.

Энэ функц нь хэд хэдэн өгөгдлийн түүвэр ( $x$  болон  $y$ ) болон таны өгөгдөлд хэрэглэх загваруудын жагсаалтыг (эдгээр нь `generate_models`-ээс олж авсан коэффициентуудын массив юм) оролтоор авна. Функц нь загвар бүрийн хувьд зураг үүсгэх ёстой. Зурагт та өгөгдлийг загвараар тодорхойлогдсон муруйтай хамт зураад,  $R^2$  утгаар тохирох чанарыг мэдээлэх ёстой. Таны график дараах форматтай тохирох ёстой:

- Өгөгдлийн цэгүүдийг дан цэнхэр цэгээр зурна.
- Таны загварыг улаан хатуу шугамаар зурна.
- Гарчиг оруулж, тэнхлэгүүдээ тэмдэглэнэ (та энэ функц нь зөвхөн  $x$ -тэнхлэг нь жил,  $y$ -тэнхлэг нь Цельсийн градус байх тохиолдолд ашиглагдана гэж үзэж болно).
  - Таны гарчиг нь загварын  $R^2$  утга, мөн загварын зэргийг агуулсан байх ёстой. Таны гарчиг нь график зурагтай харьцуулахад урт байж, хуулж буулгах үед таслагдаж магадгүй. Үүнийг засахын тулд та мөр шилжүүлэх

тэмдэгт "\n"-ийг гарчигтаа нэмж болно (жишээ нь, title = string\_a + "\n" + string\_b).

Хэрэв загвар нь шугаман муруй (өөрөөр хэлбэл, түүний зэрэг нь нэг) байвал, таны зургийн гарчиг нь энэхүү тохирсон муруйн налалтын стандарт алдааны харьцааг налалттай харьцуулсан харьцааг мөн агуулсан байх ёстой. (\*\*se\_over\_slope туслах функцийг харна уу\*\*)

Энэ харьцаа нь таны өгөгдөл дэх чиг хандлага (дээшээ/доошоо) болон тохируулах муруйг зүгээр л тохиолдлоор харах магадлалыг хэмждэг. Энэ харьцааны абсолют утга их байх тусам чиг хандлага тохиолдлоор үүссэн байх магадлал өндөр байна. Бид энэ үнэлгээний аргыг хичээл дээр үзэхгүй тул хэрэв та сонирхож байвал: Regression Slope-ийн таамаглалын тестийг шалгаж үзээрэй. Бидний тохиолдолд, хэрэв харьцааны абсолют утга 0.5-аас бага байвал чиг хандлага нь ач холбогдолтой (ө.х., тохиолдлоор үүсээгүй) байна.

## Бодлого 4: Хандлагыг Судлах

Одоо бидэнд шаардлагатай бүх бүрэлдэхүүн хэсэг байна. Бид түүхий температурын бүртгэлээс өгөгдлийн түүвэр үүсгэж, хандлагыг судалж эхлэх боломжтой.

### Бодлого 4.1 1-р сарын 10

Түүвэр авах энгийн эхний арга: бид жилээс санамсаргүйгээр нэг өдрийг (энэ тохиолдолд 1-р сарын 10) сонгож, жилүүдийн туршид температурын ямар нэгэн хандлага байгаа эсэхийг харна. Дэлхийн дулаарлаас болж энэ тодорхой огнооны температур цаг хугацаа өнгөрөх тусам нэмэгдэх ёстой гэж бид таамаглаж байна.

4.1 болон 4.II хэсгүүдийн кодоо '# Part A.4.' гэсэн тайлбарын дараа if \_\_name\_\_ == '\_\_main\_\_': доор бичнэ үү.

Эхлээд өгөгдлийн түүврүүдээ үүсгэнэ. Түүвэр бүр (өгөгдлийн цэг) нь 1961-2009 оны хоорондох жил (ө.х., СУРГАЛТЫН ЗАВСАРЛАГАН ДАХЬ жилүүд) болон тухайн жилийн Нью-Йорк хотын 1-р сарын 10-ны температур байх ёстой (үүнд туслах Climate классын функцийг харна уу!). Дараа нь, өгөгдлөө generate\_models-ийн тусламжтайгаар нэгдүгээр зэргийн полиномд тохируулж, evaluate\_models\_on\_training-ийг ашиглан регрессийн үр дүнг зурна уу. Та үүсгэсэн зургаа бичлэгтээ оруулах шаардлагатай.

### Бодлого 4.2 Жилийн Дундаж Температур

Өгөгдлийн цэгүүдийг авах өөр нэг арга замыг туршиж үзье. Дэлхийн дулаарлаас болж жил бүрийн дундаж температур цаг хугацаа өнгөрөх тусам нэмэгдэх ёстой гэж бид таамаглаж байна. Иймээс бид Нью-Йорк хотын жилийн дундаж температурын шугаман регрессийн үр дүнг зурах болно. Энэ хэсгийн кодоо '# Part A.4.' гэсэн тайлбарын дараа if \_\_name\_\_ == '\_\_main\_\_': доор бичнэ үү.

Эхлээд өгөгдлийн түүврүүдээ үүсгэнэ. Түүвэр бүр (өгөгдлийн цэг) нь 1961-2009 оны хоорондох жил (ө.х., СУРГАЛТЫН ЗАВСАРЛАГАН ДАХЬ жилүүд) болон тухайн

жилийн Нью-Йорк хотын дундаж температур байх ёстой (дахин, үүнд туслах Climate классын функцийг харна уу).

**Зөвлөмж:** өндөр жилүүдийг зөв тооцсоноо шалгаарай!

Дараа нь, өгөгдлөө `generate_models`-ийн тусламжтайгаар нэгдүгээр зэргийн полиномд тохируулж, `evaluate_models_on_training`-ийг ашиглан регрессийн үр дүнг зурна уу. Та үүсгэсэн зургаа бичлэгтээ оруулах шаардлагатай.

A4.I болон A4.II-ийн графикуудыг `ps5_writeup.pdf` баримтад оруулна уу. График бүр нь зохих тэмдэглэгээтэй тэнхлэгүүдтэй, загварын төрлөөр (жишээлбэл, шугаман, квадрат гэх мэт),  $R^2$  утга, болон стандарт алдааны харьцаагаар гарчиглагдсан байх ёстой. Та мөн `ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулах хэрэгтэй.

- Өгөгдлийг зурах тодорхой өдрийг сонгох эсвэл жилийн дунджийг тооцоолох нь бидний графикт ямар ялгаатай нөлөө үзүүлдэг вэ (ө.х.,  $R^2$  утга болон үр дүнд бий болсон муруйн тохирох байдлын хувьд)? Үр дүнг тайлбарлана уу.
- Эдгээр графикууд яагаад ийм их дуу чимээтэй байна гэж та бодож байна вэ? Аль нь илүү их дуу чимээтэй байна вэ?
- Эдгээр графикууд дэлхийн дулаарал нь температурыг нэмэгдүүлж байна гэсэн баталгааг хэрхэн дэмжиж эсвэл үгүйсгэж байна вэ? Налалт болон стандарт алдааны харьцаа нь үүний талаар бодоход тустай байж магадгүй.

## **Б хэсэг: Илүү Их Өгөгдөл Оруулах**

Нэг хотоос илүү олон хотын өгөгдлийг ашигласнаар уур амьсгалын өөрчлөлтийн талаар илүү сайн дүр зургийг олж авах боломжтой эсэхийг үзье.

Хэд хэдэн хотын жилийн дундаж температурыг авахын тулд `gen_cities_avg` функцийг хэрэгжүүлнэ үү. Энэ функцийг ашиглан 1961-2009 оны хоорондох үндэсний жилийн температурыг (ө.х., CITIES-д жагсаагдсан 21 хотын жилийн дундаж температурын дундаж) өөрийн өгөгдлийн түүвэр болгон тооцоол. Өмнөх хэсгүүдийн адилаар түүврийн x-координат нь жилийн бүхэл тоо; энэ удаад y-координат нь тухайн жилийн үндэсний жилийн температурын хөвөгч цэг юм. Таны код одоо `test_gen_cities_avg` тестийг давж өнгөрөх ёстой.

Дахин, та өгөгдлөө `generate_models`-ийн тусламжтайгаар нэгдүгээр зэргийн полиномд тохируулж, `evaluate_models_on_training`-ийг ашиглан регрессийн үр дүнг зурна уу. Энэ хэсгийн кодоо '# Part B,' гэсэн тайлбарын дараа `if __name__ == '__main__':` доор үлдээнэ үү.

Та үүсгэсэн зургаа бичлэгтээ оруулах шаардлагатай. Үр дүнг зурж, `ps5_writeup.pdf`-д оруулна уу.

`ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулна уу.

- Энэ график А хэсгийн графикуудтай хэрхэн харьцуулагдаж байна вэ (ө.х.,  $R^2$  утга, үр дүнд бий болсон муруйн тохирох байдал, болон график дэлхийн дулаарлын тухай бидний баталгааг дэмжиж эсвэл үгүйсгэж байгаа эсэхийн хувьд)? Үр дүнг тайлбарлана уу.

- Та яагаад ийм байгаа юм бол гэж бодож байна вэ?
- Хэрэв бид 3 өөр хотыг ашигласан бол үр дүн хэрхэн ялгаатай байх вэ? 100 өөр хотын хувьд яах вэ?
- Хэрэв 21 хот бүгд АНУ-ын нэг бүс нутагт (жишээлбэл, Шинэ Англи) байсан бол үр дүн хэрхэн өөрчлөгдөх байсан бэ?

## В хэсэг: 5 Жилийн Дундаж

Одоо бид 5 жилийн өгөгдлийн хөдөлгөөний дунджийг ашиглан түүврийн температурыг үүсгэх болно. Гулсах дундаж нь орон нутгийн/жилийн хэлбэлзэлээс илүү ерөнхий/дэлхийн хандлагыг тодотгох боломжийг олгодог.

`moving_average` функцийг хэрэгжүүлнэ үү. Энэ функц нь 1 хэмжээст `pylab` массив `y` болон гулсах дунджийн цонхны урт болох `window_length`-ийг оролтоор авна. Энэ нь гулсах дунджийн үр дүнг агуулсан өөр 1 хэмжээст `pylab` массивыг буцаана. Энд бид `y[i]`-ийн гулсах дунджийг `y[i-window_length+1]`-ээс `y[i]` хүртэлх дундаж гэж тодорхойлдог. Жишээлбэл, хэрэв `y=[10,20,30,40,50]` ба `window_length=3` бол гулсах дунджийн массив нь:

`[110,210+20,310+20+30,320+30+40,330+40+50]=[10,15,20,30,40]`

Зарим тохиолдолд бидэнд `window_length-1`-ээс бага өмнөх утга байх тохиолдол байдаг (массивын эхний хоёр тооцоололд байдаг шиг); ийм тохиолдолд та одоогийн утга болон байгаа өмнөх утгуудын дунджийг л авах ёстой. (Тэгэхээр массивын хоёр дахь утгын хувьд жишээлбэл, бид зөвхөн 10 ба 20-ийн дунджийг авна.)

Та одоо `test_moving_avg` тестийг давж өнгөрөх ёстой.

Энэ функцийг 1961-2009 оны үндэсний жилийн температурт ашиглан 5-ын цонхны хэмжээтэй гулсах дундаж температурыг үүсгэнэ. Дараа нь, (жил, гулсах дундаж) түүврүүдийг `generate_models`-ийн тусламжтайгаар нэгдүгээр зэргийн полиномд тохируулж, `evaluate_models_on_training`-ийг ашиглан регрессийн үр дүнг зурна уу. Энэ хэсгийн кодоо '# Part C,' гэсэн тайлбарын дараа `if __name__ == '__main__':` доор үлдээнэ үү.

Үр дүнг зурж, `ps5_writeup.pdf`-д оруулна уу.

`ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулна уу.

- Энэ график А ба Б хэсгийн графикуудтай хэрхэн харьцуулагдаж байна вэ (ө.х.,  $R^2$  утга, үр дүнд бий болсон муруйн тохирох байдал, болон график дэлхийн дулаарлын тухай бидний баталгааг дэмжиж эсвэл үгүйсгэж байгаа эсэхийн хувьд)? Үр дүнг тайлбарлана уу.
- Та яагаад ийм байгаа юм бол гэж бодож байна вэ?

## Г хэсэг : Ирээдүйг Таамаглах

Бид үнэхээр зарим хандлагыг илрүүлсэн бололтой. Одоо бид түүхэн өгөгдлөөс сурсан зүйлсдээ үндэслэн ирээдүйн температурыг таамаглаж чадах эсэхийг сонирхож байна.

## Бодлого 1: RMSE

Загваруудаа ашиглан ирээдүйн өгөгдлийн цэгүүдийг (ө.х., 1961-2009 оноос хойших өгөгдөл) таамаглахаас өмнө бид загваруудынхаа гүйцэтгэлийг хэрхэн үнэлэх талаар бодох ёстой. Бид энд  $R^2$ -ийг ашиглаж болохгүй, учир нь  $R^2$  нь туршилтын өгөгдөл дээр тодорхой утгагүй -  $R^2$  нь загварыг үүсгэхэд ашигласан өгөгдөлд загвар хэр ойр тохирч байгааг хэмждэг боловч бид 1961-2009 оны өгөгдлөөр загварыг үүсгэж, 2010-2015 оны өгөгдөл дээр шалгаж байна. Туршилтын өгөгдөл дээр загварын гүйцэтгэлийг үнэлэх нэг арга бол Үндсэн Квадрат Алдаа (RMSE) бөгөөд энэ нь таамагласан утгуудын бодит утгуудаас хэр зөрж байгааг хэмждэг.

rmse функцийг хэрэгжүүлж, загварын RMSE-г буцаана уу, энд у нь өгөгдлийн багцын бодит у-утгууд, estimated нь загвараар үнэлэгдсэн у-утгууд юм.

RMSE-г дараах байдлаар олж болно:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - e_i)^2}{n}}$$

$e_i$  нь  $i$  дэх өгөгдлийн цэгийн хувьд үнэлэгдсэн (регрессээр таамагласан) у-утга  $y_i$  нь  $i$  дэх өгөгдлийн цэгийн хувьд бодит (анхны өгөгдлөөс) у-утга  $n$  нь өгөгдлийн цэгийн тоо

Хэрэв та RMSE-ийн талаар бүрэн ойлгоогүй хэвээр байвал, түүний Wikipedia хуудас нь хэрэглээ/хэрхэн тооцоолох талаар сайн тайлбарласан байгаа.

[https://en.wikipedia.org/wiki/Root-mean-square\\_deviation](https://en.wikipedia.org/wiki/Root-mean-square_deviation)

Та одоо test\_rmse тестийг давж өнгөрөх ёстой.

Үнэлгээний хувьд та evaluate\_models\_on\_testing функцийг мөн хэрэгжүүлэх шаардлагатай. Энэ функц нь evaluate\_models\_on\_training-тэй маш төстэй ажилладаг, зөвхөн загварын таамаглалыг үнэлэхийн тулд r\_squared-ийн оронд rmse-г ашиглах ёстой. Та RMSE утгыг зургийн гарчигт мэдээлэх ёстой.

Налалтын стандарт алдааны харьцааг налалтад тооцоолох шаардлагагүй.

## Бодлого 2: Таамаглах

Одоо бидэнд туршилтын өгөгдөл дээр загваруудын гүйцэтгэлийг үнэлэх арга байна. Бид загваруудаа "ирээдүйг таамаглах"-д ашиглах болно. Загваруудынхаа таамаглалыг бодит өгөгдөлтэй харьцуулж, гүйцэтгэлийг үнэлэх хэрэгтэй тул бид 2010-2015 оны өгөгдлийг (ө.х., ТУРШИЛТЫН ЗАВСАРЛАГА) ашиглан ирээдүйг загварчлана. Бид 1961-2009 оны өгөгдлийг загвар үүсгэх "сургалтын" өгөгдлийн багц, 2010-2015 оны өгөгдлийг утгуудыг таамаглах "туршилтын" өгөгдлийн багц гэж нэрлэнэ. Энэ хэсгийн кодоо '# Part D,' гэсэн тайлбарын дараа if \_\_name\_\_ == '\_\_main\_\_': доор үлдээнэ үү.



## Бодлого 2.1 Илүү Олон Загвар Үүсгэх

Эхлээд бид таамаглахын тулд илүү олон загвар үүсгэхийг хүсэж байна. Дараах алхмуудыг хийнэ үү:

1. 1961-2009 оны үндэсний жилийн температурын 5 жилийн гулсах дунджийг сургалтын өгөгдлийн түүвэр болгон тооцоол.
2. Түүврүүдийг 1, 2, 20-р зэргийн полиномуудад тохируулна.
3. `evaluate_models_on_training`-ийг ашиглан тохирсон үр дүнг зурна уу.

Үр дүнг зурж, `ps5_writeup.pdf`-д оруулна уу.

`ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулна уу.

- Эдгээр загварууд хоорондоо хэрхэн харьцуулагдаж байна вэ?
- Алин нь хамгийн сайн  $R^2$ -тэй вэ? Яагаад?
- Аль загвар өгөгдөлд хамгийн сайн тохирч байна вэ? Яагаад?

## Бодлого 2.II Үр Дүнг Таамаглах

Одоо, зарим таамаглал хийж, таамаглалаа 2010-2015 оны бодит дундаж температуртай харьцуулъя. Дараах алхмуудыг хийнэ үү:

1. 2010-2015 оны үндэсний жилийн температурын 5 жилийн гулсах дунджийг туршилтын өгөгдлийн түүвэр болгон тооцоол.
2. Өмнөх бодлогод олж авсан загвар бүрийн хувьд (ө.х., 1961-2009 оны үндэсний жилийн температурын 5 жилийн гулсах дундажид 1, 2, 20-р зэргээр тохирсон муруйнууд), загварыг туршилтын өгөгдөлдөө (дээр тодорхойлсон) хэрэглэж, таамагласан ба бодит ажиглалтыг (ө.х., туршилтын өгөгдлийн 5 жилийн гулсах дундаж) зурна уу. Загварыг хэрэглэх болон үр дүнг зурахад `evaluate_models_on_testing`-ийг ашиглах ёстой.
3. Зэрэг = 1, 2, 20 загваруудын үр дүнд бий болсон графикуудыг зурж, `ps5_writeup.pdf`-д оруулна уу.

`ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулна уу.

- Өөр өөр загварууд хэрхэн ажилласан бэ?
- Тэдний RMSE-үүд хэрхэн харьцуулагдсан бэ?
- Аль загвар хамгийн сайн ажилласан бэ? Аль загвар хамгийн муу ажилласан бэ? Тэд Д.2.1 хэсгийнхтэй ижил байна уу? Яагаад?
- Хэрэв бид загваруудыг 22 хотын 5 жилийн гулсах дундаж ашиглахын оронд А.4.II өгөгдөл (ө.х., Нью-Йорк хотын жилийн дундаж температур) ашиглан үүсгэсэн бол 2010-2015 оны таамаглалын үр дүн хэрхэн өөрчлөгдөх байсан бэ?

## Хэсэг Е: Эрс Тэс Температурыг Загварчлах

Температурыг нэмэгдүүлэхээс гадна, дэлхийн дулаарал нь температурыг илүү эрс тэс болгодог (жишээлбэл, маш халуун эсвэл маш хүйтэн). Бид энэ нөлөөг өгөгдлийн стандарт хазайлтыг хэмжих замаар загварчилж болно гэж таамаглаж байна. Бага

стандарт хазайлт нь өгөгдөл дунджийн эргэн тойронд маш ойр байгааг илтгэнэ. Харин их стандарт хазайлт нь өгөгдөл их хэлбэлзэж байгааг илтгэнэ (ө.х., илүү эрс тэс цаг агаар). Тиймээс цаг хугацаа өнгөрөх тусам стандарт хазайлт нэмэгдэх ёстой гэж бид найдаж байна.

Бидний таамаглалыг шалгахын тулд `gen_std_devs` функцийг хэрэгжүүлнэ үү. Энэ нь `gen_cities_avg`-тэй маш төстэй байх ёстой. Энэ функц нь заасан жил бүрийн хувьд нэг утгыг агуулсан `rylab` массивыг буцаана. `years` доторх жил бүрийн хувьд таны функц дараахыг хийх шаардлагатай:

1. Тухайн жилийн өдөр бүрийн температурыг, заасан хотуудын тухайн өдрийн температурын дунджийг авч тооцоол.
2. Бүх жилийн өдөр тутмын дунджийн стандарт хазайлтыг авна.

Та одоо `test_gen_std_devs` тестийг давж өнгөрөх ёстой.

Дараа нь, функцийг туршиж үзье. Энэ хэсгийн кодоо `if __name__ == '__main__':` доорх `'# Part E,'` гэсэн тайлбарын дараа үлдээнэ үү.

1. `gen_std_devs`-ийг ашиглан 1961-2009 оны сургалтын завсарлага дахь жилүүдийн туршид 21 хотын стандарт хазайлтыг тооцоол.
2. Жилийн стандарт хазайлтууд дээр 5 жилийн гулсах дундажийг тооцоол.
3. Эцэст нь, өгөгдлөө `generate_models`-ийн тусламжтайгаар нэгдүгээр зэргийн полиномд тохируулж, `evaluate_models_on_training`-ийг ашиглан регрессийн үр дүнг зурна уу.

Үр дүнд бий болсон графикийг зурж, `ps5_writeup.pdf`-д оруулна уу.

`ps5_writeup.pdf`-д дараах асуултуудад богино хэсэгт хариулна уу.

- Үр дүн бидний баталгаатай тохирч байна уу (ө.х., температурын хэлбэлзэл эдгээр жилүүдэд нэмэгдэж байна уу)?
- Та бидний анализыг сайжруулах аргуудыг бодож олох боломжтой юу?

## Хүлээлгэн Өгөх Журам

1. Хадгалах Шийдлүүдээ `ps5.py` болон `ps5_writeup.pdf` нэрээр хадгална уу. `ps5_writeup.pdf`-ийн хувьд `.doc`, `.odt`, `.docx` гэх мэт форматаар бүү илгээгээрэй.
2. Хугацаа ба Хамтын ажиллагааны Мэдээлэл Файл бүрийн эхэнд, тайлбар хэлбэрээр, тухайн хэсгийн бодлогуудад зарцуулсан цагийн тоог (ойролцоогоор) болон хамтран ажилласан хүмүүсийн нэрийг бичээрэй. Жишээлбэл:

```
# Problem Set 5
# Name: Jane Lee
# Collaborators: John Doe
# Time:
#
... таны код энд байна ...
```

3. Хяналтын шалгалт Бодлогот даалгаврыг гүйцэтгэж дууссаны дараа хяналтын шалгалт хийнэ. Кодыг ажиллуулж, алдаагүй ажиллаж байгаа эсэхийг шалгаарай. Та шууд алдаа үүсгэдэг кодыг хэзээ ч илгээж болохгүй! Бичлэгтээ бидний хүссэн бүх зүйлийг оруулснаа шалгаарай. Ялангуяа, таны бичлэгт А.4.I-ээс нэг график, А.4.II-ээс нэг график, Б-ээс нэг, С хэсгээс нэг, Д.4.I-ээс гурван график, Д.4.II-ээс гурван график, болон Е хэсгээс нэг график багтсан байх ёстой.