

احنا هنا هنتكلم علي اللي وقفنا عندو المره اللي فاتت . بس هنعط شوية ماث ديسكربشن للي بيحصل ... واللي هنفكر فيه المره ديه هو ال algorithms نفسها .. فبالثالي هيبقا حاجه كويسه اننا بيبقي في دماغنا ال formalization of the learning process .. في notation guide هنشوفو في ال reading اللي موجود انهارد .. في سؤال الدكتور حطه ..

ليه استخدمنا الترم experience بدل ترم data .. عشان احنا مهتمين بالمصدر بتاع الداتا مش الداتا نفسها .. يعني مثلاً لو عندك عميل بيتعامل مع العالم غير انو قاعد بيقراً كتاب ..

انهارده هنبدا نتكلم علي

### Big Ideas

1. How to formalize a learning problem
2. How to learn an expert system (i.e. Decision Tree)
3. Importance of inductive bias for generalization
4. Overfitting

هنتكلم علي ازاي ن formalize a learning problem بس من الناحية ال mathematically وكمات هنتعلم expert system .. هنا احنا مش هنعمل ديزاين لل rule system بايدينا يعني .. لا احنا هنعوز نتعلم شوية ال rules اللي احنا هنمشي وراها عشان نعمل decision وده اللي هو اسمو decision tree بالنسبة لرقم 3 و 4 في الصوره اللي فوق ديه .. هنتكلم عنهم المحاضره الجايه ان شاء الله ..

سلايد رقم 14:

### FUNCTION APPROXIMATION

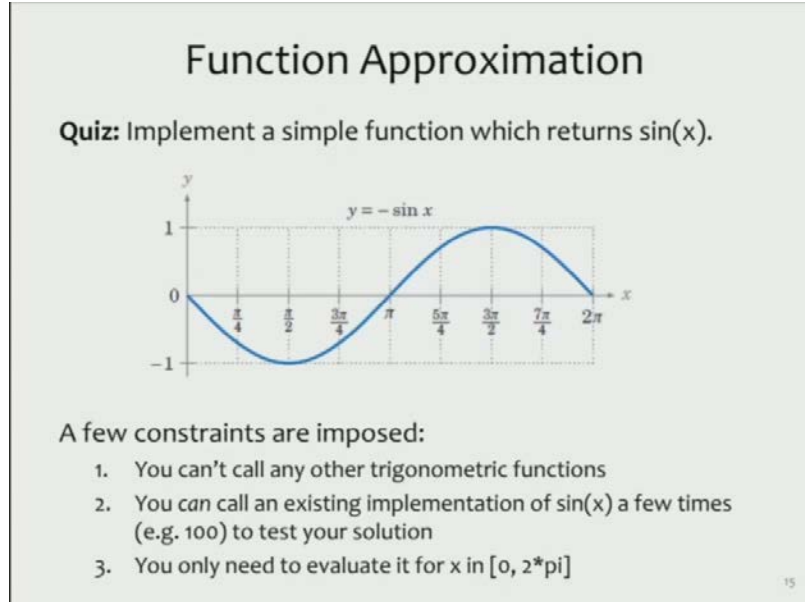
قبل ما نتعمق جامد ونبدأ الجد .. تعال نفكر في ال function approximation .. تعال نشوف كوزير الأول .. انا مثلاً لو طلبت منك انك ت implement فانكشن بسيطه بترجع  $\sin(x)$  .. بس في شوية constraints

1. انك متقدرش تنده أي trigonometric functions ثانيه
2. تقدر ت call an existing implementation of  $\sin(x)$  بضع من المرات عشان تقدر تتست الحل بتاعك
3. انت بس محتاج ت evaluate الحل بتاعك في الرينج من صفر لحد  $2\pi$

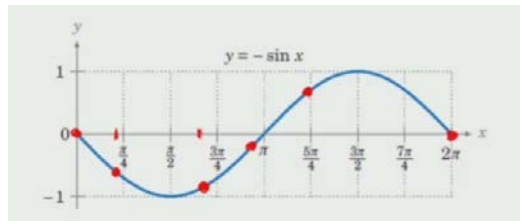
هتعمل ايه؟ " انا الحقيقه نتحت مش فاهم اعمل ايه ف ايه"

في واحد اقترح اننا ناخذ taylor series approximation وعلي حسب احنا هن expand ال series لحد فين .. ده ممكن فعلياً يدينا good mathematical approximation to sine of x بس هنا في مشكله .. انت محتاج تبقا عارف calcs .. ههه وده موضوع كبير أوي يجماعه واحنا مش ناقصين صداد ... هل في حلول ثانيه؟ بس بشرط ميكنش فيها calcs .. ما تيجي نبص علي ال constraints اللي عندنا .. في واحده الحقيقه اقترحت حل انا مفهموش .. فالدكتور بيقول ايه .. ان في ال constraints اللي موجوده هنا في الصوره الي تحت .. انك تقدر تنده existing

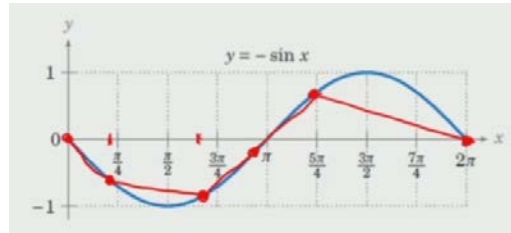
implementation of  $\sin(x)$  a few times .. فانت تقدر تتست الفانكشن 100 مره وتفضل تتابع النواتج اللي طلعتك من الفانكشن .. تعال نشوف ايه اللي هيحصل وهييqa عامل ازاي الحوار ده



فالحل كان حاجه زي الصوره ديه كدا



هنأخذ ارقام من ال  $x$  axis و نقوم جايين ال  $\sin(x)$  بتاعها .. و نقول آدي مثلاً 6 ارقام ب 6  $y$  values وبعدين هنعمل ايه .. هنوصل النقط الي احنا جبنها .  
كدا احنا بقي عندنا ايه؟ .. بقي عندنا piecewise linear function



وديه بت  $\sin(x)$  approximates the و احنا قلنا مثلاً اننا نعملها 100 مره .. فكدا تمام ممكن الحل ده فعلاً يجيب نواتج كويسه .. هل في حلول ثانيه؟ اه

بدل ما توصل النقط الي انت جبتها ب linear curve ليها متستخدامش binomial curve حاجه cubic مثلاً تقدر تضيف بيها curvature اللي بيحصل عندك ... هل في حلول ثانيه؟ اه طبعاً ... في كتير يعني بس ليه انت مهتم بالحل ده بالذات؟ .. عشان ديه طريقه فعلاً بتستخدم في ال function approximation .. بس اللي فعلاً مهم هنا .. انها ليها access to the  $\sin(x)$  function من خلال شوية نقط .. وده هو ال fundamental idea هنتحرك بيها في ال discussion بتاع انهارد .. فاحنا هنفترض اننا هنتعلم unknown function . حاجه مثلاً زي  $\sin(x)$  بس هتبقا من خلال شوية نقط ..

تعال نشوف مثال ثاني .. حاجه بتتكلم عن الداتا بتاعت medical diagnosis ... فهنا ال setting انك عارف ان الدكتور لازم لازم لازم يقرر هل المريض ده .. تعبان ولا بيستعبط فيالتالي الدكتور بيعمل ايه .. بيبص علي ال attributes بتاعت المريض عشان يعمل medical diagnosis وبعدين هيوصلو علاج لو ال diagnosis كان positive و هيقولو ان ال patient كويس جداً مفهوش حاجه لو كان ال diagnosis كان negative ..

## Medical Diagnosis

- Setting:
  - Doctor must decide whether or not patient is sick
  - Looks at attributes of a patient to make a medical diagnosis
  - (Prescribes treatment if diagnosis is positive)
- Key problem area for Machine Learning
- Potential to reshape health care

ديه فعلاً مشكله في الماشين ليرننج يعني .. احنا معندناش بالك جراوند في الطب وكدا ... فانت تقدر تروح لخبير يقولك هل ده مريض ولا لا .. و نقدر نعملهم إنترفيو هم ازاي بياخدو القرار ده ان الشخص مريض ولا لا .. و من هنا نقدر نرسم خط كدا .. ومنو نقدر نعمل ال inferences ... فدلوقت معنا سكربت ..

## Medical Diagnosis

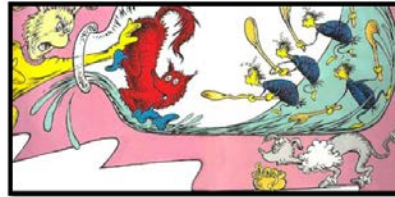
### Interview Transcript

Date: Jan. 15, 2020.

Parties: Matt Gormley and Doctor E.

Topic: Medical decision making

- Matt: Welcome. Thanks for interviewing with me today.
- Dr. E: Interviewing...?
- Matt: Yes. For the record, what type of doctor are you?
- Dr. E: Who said I'm a doctor?
- Matt: I thought when we set up this interview you said—
- Dr. E: I'm a preschooler.
- Matt: Good enough. Today, I'd like to learn how you would determine whether or not your little brother is sick given his symptoms.
- Dr. E: He's not sick.
- Matt: We haven't started yet. Now, suppose he is sneezing. Is he sick?
- Dr. E: No, that's just the sniffles.
- Matt: What if he is coughing; is he sick?
- Dr. E: No, he just has a cough.
- [Editor's note: preschoolers unilaterally agree that having the sniffles or a cough is not the same as being sick.]
- Matt: What if he's both sneezing and coughing?
- Dr. E: Then he's sick.
- Matt: Got it. What if your little brother is sneezing and coughing, plus he's a doctor.
- Dr. E: Then he's not sick.
- Matt: How do you know?
- Dr. E: Doctors don't get sick.
- Matt: What if he is not sneezing, but is coughing, and he is a fox....
- Matt: ...and the fox is in the bottle where the tweetle beetles battle with their paddles in a puddle on a noodle-eating poodle.
- Dr. E: Then he is must be a tweetle beetle noodle poodle bottled paddled muddled duddled fuddled wuddled fox in socks, sir. That means he's definitely sick.
- Matt: Got it. Can I use this conversation in my lecture?
- Dr. E: Yes



اقرأ السلايد ديه مضحكه وحلوه و لذيه

المهم قبل ما نتكلم علي ال function approximation ... هنعمل مثال كدا ع السريع .. المثال ده هيبقا عن داتا سيت حقيقه هنشتغل عليها واحنا بنفكر ازاي نوصف function approximation in a formal setting .. هنتكلم علي ال medical diagnosis .. احنا عندنا شوية expensive treatment for a disease ... فلازم الدكتور ياخذ القرار وهو عارف هو بيعمل ايه ... فالدكتور نفسو بي decide  $y \in (+, -)$

و هو بيقدر بناء علي ال attributes بتاعت المريض ... و ال attributes ديه احنا هنسميهم  $x_1, x_2, x_3, x_4, \dots, x_n$

و ال setting اللي عندنا هو ان عندنا شوية داتا بنتعلم منها .. والداتا ديه موجوده عندنا ف table .. و ال table ديه هتقولنا اهم حاجه هل المريض ده اتشخص كتعبان ولا مش تعبان ... فتعال نبص علي الصورة .. الصورة ديه بتشتغل علي السكربيت اللي في الصورة اللي فوق .. فاقراه عشان تفهم ..

Ex: Modeling Diagnosis

expensive treatment for a disease  
doctor decides  $y \in \{+, -\}$  whether the patient  
is sick based on attributes of the patient  $x_1, x_2, \dots, x_M$

sick?	sneezing?	coughing?	doctor?	Sox?
-	Y	N	N	N
-	N	Y	N	N
+	Y	Y	N	N
-	Y	Y	Y	N
+	N	Y	N	Y

فكدا احنا معانا داتا سبت نقدر نتعلم منها .. فالسؤال هنا ال learning نفسو بيبقا عامل ازاي او شكلو عامل ازاي .. احنا هنعوز نوصف الحوار ده من ال perspective بتاع function approximation واحنا هنكتب الكلام ده بطريقه تخيلنا نقدر نشوف الأمثله اللي عندنا ...

قال problem setting هنا ... هو ان عندنا set of possible inputs و set of possible outputs .. ال denotation بتاع كل واحد من اللي فاتو .. هيبقا  $X$  و  $Y$  calligraphic .. ثالث component عندنا في ال problem setting هو ال unknown target function و ده هنرمزله بالرمز  $C$  star .. و ديه هتبقا فانكشن بت  $Y$  ال set of possible outputs  $X$  ال set of possible inputs .. map .. تعال نرجع للمثال بتاعنا اللي في الصوره .. ال set of possible inputs هنا هم ال set of all possible patients in the world .. متوصفين باستخدام ال attributes بتاعة كل واحد فيهم زي الجدول كدا ... فعلي حسب المريض هيبقا عندنا قيمه مختلفه في ال  $X$  set ... فاحنا لما بنقول ان ال unknown function  $C$  star بتعمل mapping من  $X$  ل  $Y$  .. فاحنا كدا بنقول نوع الفانكشن ايه ..

## Function Approximation

### Problem Setting:

- Set of possible inputs  $\mathcal{X}$
- " " possible outputs  $\mathcal{Y}$
- **Unknown** target function  $c^*: \mathcal{X} \rightarrow \mathcal{Y}$
- Set of candidate hypotheses  $\mathcal{H} = \{h \mid h: \mathcal{X} \rightarrow \mathcal{Y}\}$

Aside:

$f(\vec{x})$

$f$

الدكتور سأل سؤال .. حد عارف ايه هو type of the function .. في ناس رفعت ايديها .. انا معرفش الحقيقه .. فالدكتور عمل نووت عشان اللي ميعرفوش زيي ..

فهو هنا في الصوره اللي تحت ديه بيقول ايه .. بيقولك افترض ان عندك  $f(x)$  و  $x$  ده عباره عن فيكتور من 3 ارقام ..  $x_1, x_2, x_3$  .. والفانكشن ديه متعرفه بالشكل  $(x_1 * x_2) + x_3$  .. "الصوره فيها حته مقطوعه أيوه" .. المهم بيقولك ان الفانكشن اللي بت map ال real valued vectors of length 3 ل real values .. فاول set هنا هو ال set of all possible real valued vectors و ال second set هي ال set of all reals

Aside: Function Types

$f(\vec{x}) = f(x_1, x_2, x_3) = (x_1, x_2)$

$f: \mathbb{R}^3 \rightarrow \mathbb{R}$

domain range

فاحنا هنستخدم ده هنقدر نتكلم علي ال actual set of candidate hypotheses .. اللي هي هتبقا ال set of possible functions اللي هن consider ك possible approximations to C star .. فديه هنعطلها رمز Calligraphic H و ده اللي هنعرفو علي أساس انو H set of functions .. بشرط ان نوع ال H هنا هو اللي بي map from X to Y ... فضروري لما بيقا عندنا ال unknown target function .. احنا هنقول ان نوع ال approximating function لازم بيقا زي ال C star .. مثلاً مثلاً يعني .. نفترض ان احنا بنحاول نبني medical diagnosis system للمشكلة اللي عندنا ... قمت انا عامل ايه .. قمت قايلك خد الفانكشن ديه اللي بتديها صورته و تقولك هل الصورة ديه فيها زباله ولا مفهش زباله زي مثلاً ال trash cleaning up robot .. فانت طبعاً هتقول ايه العته ده ... هي الفانكشن جامده و نواتجها حلوه وكل حاجه .. بس مش هي الفانكشن اللي تناسب النوع ده من المشاكل اللي بتكلم عنها . . ليه؟ .. عشان انت لما هتيجي تبيعتهلها صورة إنسان او بيانات إنسان .. مش هتاخذ منها ال feedback اللي انت عاوزو .. فعشان كذا ال type of function مهم ..

دلوقت احنا هنقول ان ال learner هنديلو الآتي ... ال training examples .. وال training examples لما هنجي نحطها كرياضه .. هتبقا set D وديه بتكون من pairs .. اللي هم x and y pairs ... فهبيقا عندنا أكثر من pair بقا جوا ال set ديه .. زي الصورة اللي تحت كذا .. وديه هتبقا ال training examples بتاعت ال unknown target function اللي احنا قلنا اسمها C star .. فاللي نقدر نقولو ان لكل i احنا عندنا  $y^i = C^*(x^i)$  في كل الكلام اللي فات ده .. ال x هنا عبارته عن vector of attributes .. واحنا بنحاول ن map ال medical diagnosis decision .. احنا شغنا ده برضو في حالة ال functional approximation example .. في حالة ال  $\sin(x)$  احنا كنا بنفكر في ال  $\sin$  function .. وهنا ال  $\sin$  function نفسها هي ال C star واحنا كنا بنعمل approximation بحاجه ثانيه .. " شوية نقط و وصل ما بينهم بخطوط مختلفه منها اللي بيغير او الكيوبيك"

Learner is Given :  
- Training examples  $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(N)}, y^{(N)})\}$   
of unknown target function  $C^*$ , where  $y^{(i)} = C^*(x^{(i)})$   
Learner Produces :  
- Hypothesis  $h \in H$  that best approximates  $C^*$

تعال نرجع ثاني لل dataset بتاعتنا اللي في الجدول اللي نسيناها ديه ... أول حاجه احنا نعملها هو اننا نضيف عمود زياده في الأول .. هنسميه i .. وده بس مجرد اننا بنعد عدد الأمثله الي عندنا .. في حاله ديه هنلاقي ان عندنا ال 1 -> 5 : .. فعدد ال N = 5 و ال M = 4 .. ال N ده هو عدد التريننج إكزامبلز و ال M ده هو عدد ال attributes .. المهم بيقولك بقا ان كل X من ال Xs اللي عندنا .. هي عبارته عن single row كامل من كل ال rows اللي عندنا .. وبيقولك ان كل واحد من ال Xs دول عندنا Y1, Y2... Y5 .. diagnosis

المهم تاخذ بالك من نقطه عشان متتعيش نفسك و تتعبنا معاك في الكلام قدام ... لما اقولك ان  $X_3^2$  ده معناه في المثال اللي قاعد فوق الإكس .. دخلنا جواه تحت .. فانت من فوق بتنزل لتحت .. فانت كذا في المثال رقم 2 .. جواه رقم 3 .. خد القرايه ديه .. فبالترتيب كذا superscript وبعدين بيجي ال subscript .. فهنا ال  $y^1$  جيه من  $C^*$  a function باستخدام ال  $X^1$  احنا مش عارفين ازاى ال C star متعرفه بس احنا عارفين انها مستخدمه عشان تطلعنا ال Y values اللي عندنا .. ففي الميديكال دايجنوزيز كان الميديكال إكسبيرت هو اللي اداها ال C star فده دكتور هو اللي عمل الدايجنوزيز اللي في الجدول ديه .. و أياً كان ال decision process اللي هم بيستخدموها عشان يعملو ال diagnosis ديه .. هي ديه ال C star فال learner هنا هو ال learner will produce a single hypothesis .. فهم هي Calligraphic H في ال set H ال produce some hypothesis that is some H in the set و نأمل ان اللي هيطلع هنا هو اللي أحسن تقريب لل C star

i	sick?	sneezing?	coughing?	doctor?	fox?	
1	-	Y	N	N	N	$x^{(1)}$
2	-	N	Y	N	N	$x^{(2)}$
3	+	Y	Y	N	N	$x^{(3)}$
4	-	Y	Y	Y	N	$x^{(4)}$
5	+	N	Y	N	Y	$x^{(5)}$

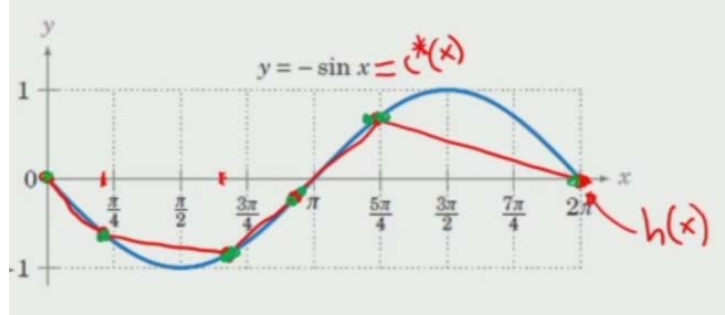


تعال نرجع للمثال بتاع ال  $\sin(x)$  ثاني ... بس المره ديه هنحط ال notation بتاعنا

$$Y = \sin(x) = C^*(x) = \text{unknown target function}$$

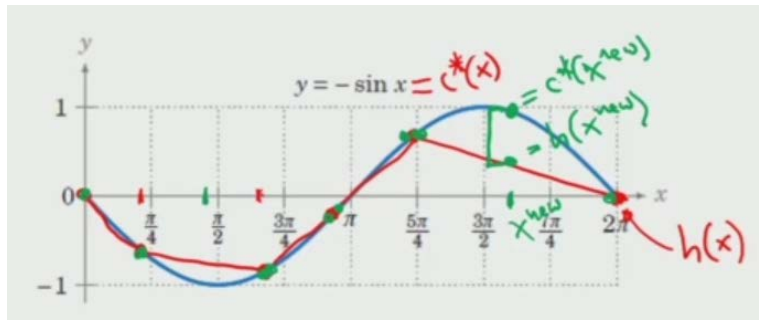
و ال  $h(x)$  piecewise function هي ال

السؤال هنا .. هل ال  $h(x)$  هي good approximation for  $C^*(x)$  .. حد قال لأ .. سالو ليه؟ .. قالو عشان لو احنا كنا خدنا شوية نقط كمان كان زمان الدنيا أحسن .. فالدكتور قام عامل كذا عشان يعني "يرخّم" عليه



بس قالو كذا متحسنتش .. فقام سأل السؤال ... هو احنا بنقول ازاى ان الفانكشن اتحسنت ولا لا .. فسأل ازاى هنعقيس الكوالتي بتاعت الحاجه ديه .. حد رد عليه وقالو علي اقتراح وقام هو مكمل عليه . في الآخر .. خد شوية نقط ثانيه مختلفه عن اللي معاك .. ونقول إن القيمه الحقيقيه بتاعتها هي ال  $C^8(X_{new})$

.. وبعدين نقوم جايبين القيمه بتاعتنا من علي الجراف اللي احنا عملناه اللي هو اسمو  $h(x_{new})$  و نقوم قايلين السؤال الجامد جداً ده .. how much a gap is there between the top and the bottom point .. لو الجاب كانت كبيره .. ساعتها الأبروكسميشن بتاعك مش كويس ... لو الجاب صغيره .. يبقى انت عال العال ..



اللي عاوزين نعملو دلوقت اننا هنعرف حاجه اسمها ال loss function .. و هنعوز نجابو علي السؤال بتاع ? how to evaluate loss .. فال function هتبقا function نسميها L وبتاخذ قيمتين x و y و تجيب قيمه Real value و تقيس

How bad a prediction is  $y^{\wedge} = h(x)$  are compared to  $C^*(x)$

ودلوقت شغلانة ال practitioner هو انو يختار ايه هي ال loss function اللي هنستخدمها ..

خلينا نفترض إن عندنا regression problem و في المشكله ديه احنا عندنا ال y values بتيجي من set of real numbers .. و في الحاله ديه .. مكن نعرف loss function بتاخذ قيمة الواي و قيمة ال y predicted و نقوم تحسب الفرق ما بينهم .. وبعدين تربع الفرق ده .. فده اسمو squared loss

دلوقت احنا هنقول ان مشكلتنا هي عبارته عن classification problem ف هنا ال y is discrete . فهنقول ان ال  $loss(y, yhat)$  .. هنرجع قيمتين .. يا صفر يا واحد .. الصفر في حالة ان ال  $y = yhat$  .. فهل ده منطقي في حالة الكلاسيفيكيشن؟ لو جينا نبص علي مثال واحد .. ده هيقولنا ده ان لو كان زي بعضو بيقا عندي zero loss لو كان incorrect بيقا عندي  $loss = 1$  وكل ما كان اعلي كل ما كان اسوا .. لو خدنا ال average بتاع ال loss .. فده بيودينا لناحية اننا نعرف حاجه اسمها error rate .. فالإيرور هيتحسب across N examples

Practitioner Chooses...

Ex: Regression  $y \in \mathbb{R}$   $l(y, \hat{y}) = (y - \hat{y})^2$  "squared loss"

Classification  $y$  is discrete  $l(y, \hat{y}) = \begin{cases} 0 & \text{if } y = \hat{y} \\ 1 & \text{otherwise} \end{cases}$  "zero-one loss"

$= \mathbb{1}(y \neq \hat{y})$  indicator function

$$= \mathbb{I}(y \neq \hat{y}) \quad \text{indicator}$$

Def: Error Rate for 0/1 loss

Let  $D$  be a dataset and  $h(x)$  be a hypothesis

$$\text{error}(D, h) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(y^{(i)} \neq h(x^{(i)}))$$

so  $\text{error}(D, h) \in [0, 1]$  and can use this to measure for a "best" approximation.

Another dataset  $D^{\text{test}} = \{(x^{(1)}, y^{(1)}), \dots, (x^{(N')}, y^{(N')})\}$

What is our average loss on  $D^{\text{test}}$ ? e.g.  $\text{error}(D^{\text{test}}, h)$

تعال نفترض ان في داتا سيت تانيه .. ونسميها .. D test .. زي الاولانيه هتتكون من شوية إكسات و شوية وايات .. فهنا ايه هي الأفريديج لوسس بتاعت ال تستنج داتا سيت .. فده هيفتحلنا مساحه بقولنا قد ايه ال approximation is good

Learner is Given:

- Training examples  $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(N)}, y^{(N)})\}$   
of unknown target function  $c^*$ , where  $y^{(i)} = c^*(x^{(i)})$

Learner Produces:

- Hypothesis  $h \in \mathcal{H}$  that best approximates  $c^*$

To Evaluate:

- Loss Function  $\ell: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$  measures how "bad" predictions  $\hat{y} = h(\tilde{x})$  are compared to  $c^*(\tilde{x})$

Practitioner Chooses...

Ex: Regression  $y \in \mathbb{R}$   $\ell(y, \hat{y}) = (y - \hat{y})^2$  "squared loss"

Classification  $y$  is discrete  $\ell(y, \hat{y}) = \begin{cases} 0 & \text{if } y = \hat{y} \\ 1 & \text{otherwise} \end{cases}$  "zero-one loss"  
 $= \mathbb{I}(y \neq \hat{y})$  indicator function

Def: Error Rate for 0/1 loss

Let  $D$  be a dataset and  $h(x)$  be a hypothesis

$$\text{error}(D, h) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(y^{(i)} \neq h(x^{(i)}))$$

so  $\text{error}(D, h) \in [0, 1]$  and can use this to measure for a "best" approximation.

- Another dataset  $D^{\text{test}} = \{(x^{(1)}, y^{(1)}), \dots, (x^{(N')}, y^{(N')})\}$

- What is our average loss on  $D^{\text{test}}$ ? e.g.  $\text{error}(D^{\text{test}}, h)$

كدا انت بقي عندك التستنج وعندك التريننج ... فانت لما هتعمل التريننج هييجبك single h .. تقوم رايح بت evaluate علي ال testing data ..

تعال نتكلم علي algorithm نقدر نستخدمو عشان يعمل الحركه ديه .. اول واحد هو algorithm 0 ... ليه صفر؟ .. عشان ده بدايه يعني .. هنسميه memorizer وهيتكون من 2 فانكشنز ..

Def train (DataSet D):

Store DataSet D

// Prediction function and takes a single new example

Def h(x):

If there exists some  $X^i$  in our DataSet D such that  $X^i = X$  that we passed in:

Then return  $y^i$

Else

Return y randomly

تعال نفترض اننا هنشتغل بالألجورزم ده .. هل ال memorization a form of learning?

الاجابه لا ... ازاي هتطلع من داتا انت عارفها لداتا انت مش عارفها .. هو بتتعلم تمام .. بس مش هتتعرف ت generalize ...

تعال نفكر في عدد ال patient types ...

How many patients for the example above? .. what is the size of set of possible patients?

2 \* 2 \* 2 \* 2 = 16 ليه؟ .. 2 أوبشنز لكل فيتشر من الفيتشرز اللي عندك فوق .. 2 للعطس 2 للكه 2 للفلوكس 2 دكتور ..

طب ايه اللي هيحصل لو قلنا ان عندنا أكثر من attribute او أكثر من feature يعني .. يعني مثلاً هنضيف age – preschooler – history – history illness

فاحنا عندنا binary attributes as example .. لو عندنا 10 attributes .. في الحاله ديه هيبقا حجم ال X هيبقا 2 أس 10 .. عشان هي باينري ..

بس غالباً هيبقي عندنا 100 attributes .. هيبقا عندنا 2 أس 100 .. ده عدد المرضي اللي ممكن يكونوا موجودين .. فكدا كثير جداً الحقيقه ....

بس الدكاتره الحقيقه بقا بتستخدم more attributes .. فكدا في مشكله في الألجورزم ده ..

تعال نشوف ألجورزم 1...

Algorithm 1: Majority vote classifier

Like the memorizer, define 2 functions

Def train (DataSet D):

Store a single vote (v) which is the majority vote on DataSet D

This function returns the actual class y from the set of possible Ys that appears most often in the DataSet D

Def h(x):

Return Vote you already stored

تعال نفترض ان الداتا سبت بتاعتنا .. ايه اللي هيرجع من الألجورزم ده لما نقول sneezing coughing fox

.. هيرجع سالب .. ليه؟ ... لأنو هيبص علي ال labels فهيقوم عامل ايه . يلاقي في 3 سالب و 2 موجب .. هيرجع سالب .. فبغض النظر عن الفيتشرز اللي موجوده في أي مثال عندنا .. اللي هيرجع دائماً سالب ..

لما هتيجي تفكر هل ده ألجورزم قادر يتعلم؟ .. آه .. افترض ان عندك مرض نادر جداً جداً .. لو انت جمعت أوقات من المرضي فالألجورزم ده .. هيرجع بشكل صحيح .. انو لو رجع سالب لكل المرضي اللي هو شايفهم هيبقا صح معظم الوقت .. والايرور قليل .. بس ده هيعمل ميديكال ديسيजनز .. هيقول لكل الناس .... انتو معندكوش مشاكل ... فده موقف انت عندك ال loss function مش بت match what you have .. الصفر و الواحد losses هتقول علي ال rare disease ان ال majority vote ليه برفورمنس فشيخ .. بس مش ده اللي انت عاوزو .. انت هتحتاج تعدل ال loss function



## Majority Vote Classifier Example

### Dataset:

Output Y, Attributes A and B

Y	A	B
-	1	0
-	1	0
+	1	0
+	1	0
+	1	1
+	1	1
+	1	1
+	1	1

### In-Class Exercise

What is the **training error** (i.e. error rate on the training data) of the **majority vote classifier** on this dataset?

Choose one of:  
 $\{0/8, 1/8, 2/8, \dots, 8/8\}$

23

فالإيرور هنا هو 2 من 8 ... لأنو هيقول إن كل حاجه لازم تبقي بوزيتيف .. فهتلاقي ان 2 بس هو النيجاتيف .. فكدا الألبورزم ده naive .. تعال نشوف واحد احسن ..

### Algorithm 2: Decision stump

Def train (DataSet D):

1. Pick an attribute,  $m$  i.e. = 1
2. Divide the DataSet D on attribute  $m$ 
  - a. You have 2 DataSets (Partitioned DataSet)
    - i.  $D_0$  -> all  $x$  and  $y$  pairs have  $x = 0$
    - ii.  $D_1$  -> all  $x$  and  $y$  pairs have  $x = 1$
3. Take 2 votes:
  - a.  $V_0$  = majority vote for  $D_0$
  - b.  $V_1$  = majority vote for  $D_1$

Def  $h(x)$ :

If  $x = 0$ : return  $V_0$

Else: return  $V_1$

### Algorithm 2: Decision Stump

def train (D):

① pick an attribute,  $m$

② divide dataset D on  $m$

$$D^{(0)} = \{(x, y) \in D \mid x_m = 0\}$$

$$D^{(1)} = \{(x, y) \in D \mid x_m = 1\}$$

③ two votes

$$v^{(0)} = \text{majority\_vote}(D^{(0)})$$

$$v^{(1)} = \text{majority\_vote}(D^{(1)})$$

def  $h(x)$ :

if  $x_m = 0$ :

return  $v^{(0)}$

else ( $x_m = 1$ ):

return  $v^{(1)}$

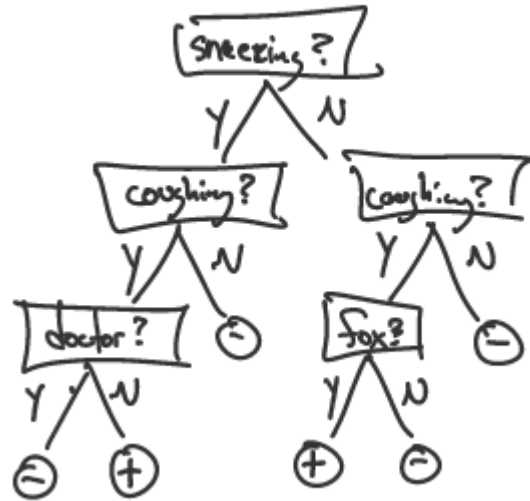


فتعال نبص علي نفس الداتا سبت .. نبص علي ال sneezing .. في عندك 3 دواير باللون السماوي او الأزرق .. فهتقسم الفوت بتاعك .. لما الواي بتساوي واحد .. ده النيجاتيف .. لما الواي بتساوي صفر ده بوزيتيف ...

### Algorithm3: Decision Tree

شوية أسئلة ومعهاا شجره بقا ..

#### Algo 3: Decision Tree



ده كذا اسمو rule based system ... شوية اسئلة تقدر نسألها.. الالجورزم رقم 2 .. كان بيعمل حاجه شبه كذا ... بس كان ناقص سؤال واحد بس 00 ازاى نختار ال attribute اللي هنقسم بناء عليه

هنشوف الحاجات ديه في الhomework