به نام خداوند بخشنده مهربان

جبر خطی

دانشکده مهندسی کامپیوتر

حمیدرضا ربیعی، مریم رمضانی

بهار ۱۴۰۲

فضای برداری و دستگاه معادلات | **تمرین اول** | تاریخ انتشار: ۲۵ بهمن ۱۴۰۱

---

۱. پرسش‌های خود در مورد این تمرین را در سامانه کوئرا مطرح کنید.

۲. **سیاست ارسال با تاخیر پاسخ:** شما در مجموع در طول نیم‌سال می‌توانید تا سقف ۱۵ روز تمارین خود را با تاخیر ارسال کنید. پس از آن به ازای هر ساعت تاخیر ۰.۵ درصد نمره کسر خواهد شد. مقیاس محاسبه تاخیر ساعت است و تاخیر شما رو به بالا گرد خواهد شد تا به مقیاس ساعت درآید. به عنوان مثال ۲ ساعت و ۱۰ دقیقه تاخیر معادل ۳ ساعت در نظر گرفته خواهد شد. تمارین تئوری و عملی در محاسبه تاخیرها دو تمرین جداگانه در نظر گرفته خواهند شد. یعنی ۲ روز تاخیر بر تمرین تئوری و ۱ روز تاخیر بر تمرین عملی در مجموع برابر ۳ روز در نظر گرفته خواهد شد. در نهایت اگر بیش از ۱۵ روز تاخیر داشته باشید، تاخیر مجاز شما به نحو بهینه‌ای توزیع خواهد شد که کمترین میزان نمره را از دست بدهید.

۳. **سیاست مشارکت دانشجویان در حل تمارین:** دانشجویان می‌توانند در حل تمارین برای رفع ابهام و یا بدست آوردن ایده کلی با یک دیگر مشارکت و هم‌فکری کنند. این کار مورد تایید و تشویق تیم ارائه درس می‌باشد چرا که هم‌فکری و کار گروهی می‌تواند موجب تقویت یادگیری شود. اما بدست آوردن جزئیات راه‌حل و نگارش پاسخ باید تماما منحصرا توسط خود دانشجو انجام شود. بیشینه اندازه مجاز گروهی که در آن می‌توانید برای حل تمرین هم‌فکری کنید برابر ۴ نفر است. رعایت آداب‌نامه‌ی انجام تمرین‌های درسی در انجام تمارین الزامی است. در صورت مشاهده تخلف مطابق قوانین دانشکده و دانشگاه برخورد خواهد شد.

سوالات تئوری (۱۴۰ نمره)

تاریخ تحویل: ۶ اسفند ۱۴۰۱

پرسش ۱ (۲۸ نمره) برای هرکدام از $(V, \odot, \oplus)$ تعریف شده مشخص کنید که آیا یک فضای برداری داریم یا خیر. ($c \in \mathbb{R}$)

(آ) (۹ نمره) $V = P_2$، تمام چندجمله‌ای‌ها از درجه حداکثر ۲ میباشد ، همینطور $p(t) \oplus q(t) = p'(t)q'(t)$، $c \odot p(t) = cp(t)$.

(ب) (۹ نمره) $V = \{v_1, v_2\}$ و داریم $v_1 \oplus v_1 = v_2 \oplus v_2 = v_1$ و $v_1 \oplus v_2 = v_2 \oplus v_1 = v_2$ و در آخر $c \odot v_2 = c \odot v_1 = v_1$. ($v_1, v_2 \in \mathbb{R}^n$)

(ج) (۱۰ نمره) $V = \mathbb{R}^2$ و $c \odot (x, y) = (cx, |c|y)$ و $(x_1, y_1) \oplus (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$.

پاسخ

(آ) در این تعریف از جمع، عضو خنثی وجود ندارد. به طور مثال فرض کنید میخواهیم برای توابع $x^2 + x$, x عضو خنثی پیدا کنیم. داریم:

$$x = q(x), (2x + 1)q(x) = x^2 + x$$

واضح است که عضو خنثی برای این دو عضو یکی نیست. (تابع دوم عضو خنثی ندارد)

(ب) در ضرب تعریف شده عنصر خنثی وجود ندارد. به این معنی که برای عضو v_2 هیچ ضریبی وجود ندارد که داشته باشیم $cv_2 = v_2$.

(ج) در این تعریف ضرب، خاصیت اشتراک پذیری ندارد. داریم:

$$(1 - 1)u = 0, 1u + (-1)u = (0, 2y) \neq 0$$

پرسش ۲ (۲۸ نمره) به سوالات زیر پاسخ دهید.

(آ) (۷ نمره) فضای برداری V را به صورت تمام توابع مشتق پذیر بر روی $\mathbb{R}$ تعریف میکنیم. همینطور S را بر روی V به صورت مجموعه تمام توابع چندجمله‌ای تعریف میکنیم. نشان دهید که $f(x) = \sin(x)$ در $\text{Span}(S)$ قرار ندارد.

(ب) (۷ نمره) فضای برداری V را مانند قسمت قبل درنظر بگیرید. نشان دهید تمام توابعی که $f' + 2f = 0$ برای آن‌ها برقرار باشد، یک زیرفضا را میسازند.

(ج) (۷ نمره) توابع p_1, p_2, p_3 را به صورت $p_1(x) = 2 + x + 4x^2$، $p_2(x) = 1 - x + 3x^2$ و $p_3(x) = 3 + 2x + 5x^2$ تعریف میکنیم. توابع $q_1(x) = -9 - 7x - 15x^2$ و $q_2(x) = 0$ را به صورت ترکیبی خطی از p_1, p_2, p_3 پیدا کنید.

(د) (۷ نمره) داریم $f(x) = \sin^2(x)$ و $g(x) = \cos^2(x)$، مشخص کنید که توابع 0 , 1 , $3 + x^2$, $\cos(2x)$ در $\text{Span}(f, g)$ قرار دارند یا خیر.

پاسخ

(آ) فرض کنید اینکار ممکن باشد یعنی چندجمله‌ای درجه m وجود داشته باشد به طوری که به توان تابع $\sin(x)$ را به صورت $\sin(x) = \mathcal{P}_m$ نوشت. میتوان دید که $\sin(x) - \mathcal{P}_m \sim \mathcal{O}(m + 2)$. بنابراین اینکار امکان ندارد.

(ب) لازم است ۳ خاصیت زیرفضا را برای این مجموعه از توابع اثبات کنیم. مجموعه تمام این توابع را U در نظر بگیرید.

$$f = ۰ \to f' + ۲f = ۰ \to f \in U$$

$$f, g \in U \to f' + ۲f = ۰, g' + ۲g = ۰ \to (f+g)' + ۲(f+g) = ۰ \to f + g \in U$$
$$f \in U \to f' + ۲f = ۰ \to (\alpha f)' + ۲\alpha f = ۰ \to \alpha f \in U$$

(ج) نوشتن تابع $q_۲$ که بدیهی میباشد حال لازم است تابع $q_۱$ را بنویسیم.

$$\begin{bmatrix} ۴ & ۳ & ۵ \\ ۱ & -۱ & ۲ \\ ۲ & ۱ & ۳ \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} -۱۵ \\ -۷ \\ -۹ \end{bmatrix}$$

با درآوردن این ماتریس به فرم سطری پلکانی میبینیم که ۳ عنصر محوری دارد و بنابراین معادلات جواب دارند و جواب یکتا وجود دارد.

(د) واضح است که داریم:

$$\cos^۲(x) - \sin^۲(x) = \cos(۲x)$$
$$۰ \times \cos^۲(x) + ۰ \times \sin^۲(x) = ۰$$
$$\cos^۲(x) + \sin^۲(x) = ۱$$

حال نشان میدهیم که ساختن تابع $x^۲ + ۳$ ممکن نمیباشد. در ابتدا فرض خلف میکنیم و فرض میکنیم که ضرایب a, b وجود دارند که این تابع را بسازند.

$$a \sin^۲(x) + b \cos^۲(x) = ۳ + x^۲$$

$$\to a = ۳ + \frac{\pi}{۲}^۲, a = ۳ + \frac{۵\pi}{۲}^۲$$

بنابراین این مقدار از a وجود ندارد و نمیتوان تابع $x^۲ + ۳$ را ساخت.

پرسش ۳ (۲۸ نمره) به پرسش‌های زیر پاسخ دهید.

(آ) (۱۴ نمره) V یک فضای برداری و ϕ یک مجموعه از زیرفضاهای V میباشد. مجموعه $\cap_\phi$ را به صورت زیر تعریف میکنیم.

$$\cap_\phi = \{v \mid v \in V \;,\; \forall_{U \in \phi}\; v \in U\}$$

نشان دهید که $\cap_\phi$ خود یک زیرفضا است.

(ب) (۱۴ نمره) فرض کنید $W_۱, \ldots, W_n$ زیرفضاهایی از V باشند و ϕ مجموعه‌ای که از تمام زیرفضاهایی باشد که شامل تمام W_i ها هستند.

$$\phi = \{U \mid \forall_{W_i} W_i \subseteq U\}$$

نشان دهید که داریم:

$$\cap_\phi = W_۱ + \cdots + W_n$$

پاسخ

(آ) دقت کنید که مجموعه $\cap_\phi$ اشتراک تمام زیرفضاهایی است که در ϕ قرار دارند. بنابراین لازم است نشان دهیم که اشتراک چند زیرفضا خود یک زیرفضا است.

$$\begin{cases} ۰ \in U_۱ \\ \vdots \\ ۰ \in U_n \end{cases} \to ۰ \in \bigcap_i U_i \to ۰ \in \cap_\phi$$

$$u \in \cap_\phi \to \begin{cases} u \in U_۱ \\ \vdots \\ u \in U_n \end{cases} \to \begin{cases} \alpha u \in U_۱ \\ \vdots \\ \alpha u \in U_n \end{cases} \to \alpha u \in \cap_\phi$$

$$u, v \in \cap_\phi \to \begin{cases} u, v \in U_۱ \\ \vdots \\ u, v \in U_n \end{cases} \to \begin{cases} u + v \in U_۱ \\ \vdots \\ u + v \in U_n \end{cases} \to u + v \in \cap_\phi$$

پس $\cap_\phi$ خود یک زیرفضا میباشد.

(ب) فرض کنید U یک زیرفضا در ϕ باشد. بنابراین داریم:

$$\forall_i W_i \subseteq U \to W_۱ + \cdots + W_n \subseteq U$$

این نتیجه از آنجا حاصل میشود که U خود یک زیرفضا است و هر ترکیب خطی از اعضای آن در خودش قرار دارد. بنابراین هر جمعی از اعضای W ها در خودش قرار دارد. حال میدانیم:

$$\forall_{U\in\phi} W_۱ + \cdots + W_n \subseteq U \to W_۱ + \cdots + W_n \subseteq \cap_\phi$$

از طرفی دیگر میدانیم که خود $W_۱ + \cdots + W_n$ یکی از اعضای ϕ است چرا که داریم $\forall_i W_i \subseteq W_۱ + \cdots + W_n$. بنابراین داریم:

$$\cap_\phi \subseteq W_۱ + \cdots + W_n$$

با استفاده از دو رابطه زیرمجموعه‌ای بدست آمده داریم:

$$\cap_\phi \subseteq W_۱ + \cdots + W_n, W_۱ + \cdots + W_n \subseteq \cap_\phi$$

$$\to W_۱ + \cdots + W_n = \cap_\phi$$

پرسش ۴ (۲۸ نمره)

(آ) (۷ نمره) W زیرفضایی در فضای برداری V میباشد. نشان دهید زیرفضای W' وجود دارد به طوری که $W + W' = V$ و همینطور $W \cap W' = \{۰\}$.

(ب) (۷ نمره) زیرمجموعه ناتهی U از $\mathbb{R}^۲$ را در نظر بگیرید که تحت عمل جمع و تفریق بسته میباشد. نشان دهید که U میتواند یک زیرفضا نباشد.

(ج) (۷ نمره) S زیرفضایی از V میباشد همینطور مجموعه‌ای از اعضای V را به صورت $B = \{b_۱, \ldots, b_k\}$ را در نظر بگیرید که $\forall_{v\in V}\ v \in \text{Span}(B)$. درستی عبارت زیر را بسنجید.

$$S = \bigoplus_{i=۱}^{k} (S \cap \text{Span}(b_i))$$

(د) (۷ نمره) برای زیرفضاهای $W_۱, \ldots, W_k$ در V به طوری که $W_i \neq V$ نشان دهید که عضوی در V وجود دارد که در هیچ یک از W_i ها نمیباشد.

پاسخ

(آ) مجموعه W' را به صورت مجموعه‌ای از تمامی عضوهایی از V تعریف میکنیم که به تمامی اعضای W عمود هستند. واضح است که W' خود یک زیرفضا میباشد و تنها اشتراکش با W عضو صفر است:

$$\forall_{u\in W} ۰^T u = ۰ \to ۰ \in W'$$

$$u \in W' \to \forall_{v\in W} u^T v = ۰ \to (\alpha u)^T v = ۰ \to \alpha u \in W'$$

$$u', v' \in W' \to \forall_{u\in W} u^T v' = ۰, u^T u' = ۰ \to u^T(u' + v') = ۰ \to u' + v' \in W'$$

بنابراین W' یک زیرفضا است و داریم:

$$u \in W \cap W' \to u^T u = ۰ \to u = ۰$$

بنابراین تنها اشتراک این دو زیرفضا عضو صفر میباشد. حال لازم است ثابت کنیم که هر عضوی در V را میتوان به صورت جمعی از اعضای W, W' نوشت. به ازای هر عضو در V نزدیک‌ترین عضو آن را در W در نظر بگیرید. میتوان نشان داد که $v - W^*$ در W' قرار دارد به طوری که v عضو موردنظر در V است و w^* نزدیک‌ترین عضو موردنظر است. دقت کنید که نزدیک‌ترین عضو یکتا است و همچنین وجود این عضو در W' به این دلیل است که اگر اینطور نباشد عضو نزدیک‌تری به v یافت خواهد شد. بنابراین مجموعه W' که ساختیم تمام خواص خواسته شده را دارد.

(ب) مجموعه زیر یک مثال نقض برای این حکم است.

$$\{\begin{bmatrix} x \\ x \end{bmatrix} \mid x \in \mathbb{Z}\}$$

(ج) مجموعه‌های زیر مثال نقضی برای این حکم میباشند.

$$S = \{\alpha \begin{bmatrix} ۱ \\ ۱ \end{bmatrix} \mid \alpha \in \mathbb{R}\}$$

$$B = \{\begin{bmatrix} ۱ \\ ۰ \end{bmatrix}, \begin{bmatrix} ۰ \\ ۱ \end{bmatrix}\}$$

داریم:

$$\bigoplus_{i=۱}^{k} (S \cap \text{Span}(b_i)) = \{۰\} + \{۰\} = \{۰\} \neq S$$

به عنوان تمرین حالتی از این مسئله را حل کنید که داریم:

$$S \cap b_i \neq \{۰\}$$

و جواب را مقایسه کنید.

(د) لم: اگر اجتماع ۲ زیرفضا $W_۱, W_۲$ یک زیرفضا باشد، یکی از آنها زیرمجموعه دیگری است.

اثبات: فرض کنید اینطور نباشد بنابراین $\exists x \in W_۱, x \notin W_۲$ و همچنین $\exists y \in W_۲, y \notin W_۱$. حال میدانیم اجتماع آنها یک زیرفضا است پس $x + y$ هم در اجتماع آنها قرار دارد. بدون از دست رفتن کلیت، فرض کنید که $x + y \in W_۱$. بنابراین داریم $y \in W_۱$ که این خلاف فرض است و بنابراین یکی از این زیرفضاها در دیگری قرار دارد.

لم: اگر اجتماع زیرفضا های $W_۱, \ldots W_k$ خود یک زیرفضا باشد آنگاه وجود دارد i به طوری که $W_i = \bigcup_j W_j$.

اثبات: زیرفضای اجتماع آنها را با W' نشان میدهیم. حال فرض کنید داشته باشیم $W_۱ = W'$. بدیهتا در این حالت حکم اثبات شده است. حال لازم است حالتی را بررسی کنیم که $W_۱ \neq W'$. عضوی از W' که در $W_۱$ قرار ندارد را y مینامیم. فرض کنید داشته باشیم $W_۱ \not\subseteq \bigcup_{j=۲}^{k} W_j$. بنابراین فرض کردیم که نه $W_۱$ خود برابر کل فضا است و نه در اجتماع زیرفضاهای دیگر قرار دارد. حال عضوی از $W_۱$ که در اجتماع زیرفضاهای دیگر نیست را x مینامیم. خواهیم دید که در این حالت به تناقض خواهیم خورد. اعضای زیر را در نظر بگیرید (هرکدام از α_i ها متفاوت هستند.

$$\alpha_۱ y + x, \ldots, \alpha_k y + x$$

میدانیم x در $W_۱$ قرار دارد که خود زیرمجموعهای از W' است. بنابراین تمامی اعضا ذکر شده در W' قرار دارد و هرکدام باید در یکی از $W_۲, \ldots, W_k$ آمده باشد چرا که میدانیم هیچکدام از آنها در $W_۱$ نیستند (چون در اینصورت y هم در $W_۱$ خواهد بود که خلاف فرض است.) بنابراین k نقطه داریم که هرکدام در حداقل یکی از $k - ۱$ زیرفضا دیگر قرار دارند. بنابراین یکی از آنها وجود دارد که ۲ عضو از این اعضا نام برده شده را دارد. پس میتوانیم نتیجه بگیریم که x درون یکی از $k - ۱$ زیرفضا دیگر قرار دارد و بنابراین فرض $W_۱ \not\subseteq \bigcup_{j=۲}^{k} W_j$ رد میشود و داریم

$$W_۱ \subseteq \bigcup_{j=۲}^{k} W_j$$

و

$$\bigcup_{j=۱}^{k} W_j = \bigcup_{j=۲}^{k} W_j = W'$$

بنابراین میتوانیم مسئله را بدون $W_۱$ حل کنیم. اینکار را به صورت استقرایی انجام میدهیم تا به $k = ۲$ برسیم که در لم اول اثبات کردیم.

حال از این لم استفاده میکنیم تا سوال را حل کنیم. فرض کنید به ازای هر عضو درون V، این عضو درون یکی از W_i ها آمده باشد بنابراین اجتماع آنها باید برابر با V باشد که خود یک زیرفضا است. حال میدانیم اجتماع k زیرفضا خود یک زیرفضا شده است بنابراین باید یکی از آنها وجود داشته باشد که برابر با V باشد که این خلاف فرض اولیه است. بنابراین عضوی در V وجود دارد که در هیچ یک از W_i ها وجود ندارد.

پرسش ۵ (۲۸ نمره) برای تمامی مقادیر k مشخص کنید که چه زمانی دستگاه معادلات زیر جواب ندارد، جواب یکتا دارد و یا بینهایت جواب دارد.

$$x + ۵y - ۳z = ۲$$

$$-۲x - ۷y + ۳z = -۵$$

$$-x - ۵y + (k^۲ - ۶)z = k + ۱$$

پاسخ معادله فوق را به فرم ماتریسی بازنویسی میکنیم.

$$\begin{bmatrix} ۱ & ۵ & -۳ \\ -۲ & -۷ & ۳ \\ -۱ & -۵ & k^۲ - ۶ \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} ۲ \\ -۵ \\ k + ۱ \end{bmatrix}$$

حال معادل کاهش یافته سطری پلکانی این ماتریس را پیدا میکنیم.

$$\begin{bmatrix} ۱ & ۵ & -۳ \\ ۰ & ۳ & -۳ \\ ۰ & ۰ & k^۲ - ۹ \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} ۲ \\ -۱ \\ k + ۳ \end{bmatrix}$$

به ازای تمام مقادیر حقیقی k به غیر از ۳ و -۳ ماتریس ضرایب سه عنصر محوری دارد و در نتیجه معادله جواب یکتا دارد. در حالت $k = ۳$ یکی از معادلات به فرم $۰z = ۶$ بدست می‌آید که جواب ندارد. در حالت $k = -۳$ معادله آخر جواب خواهد داشت و z یک متغیر آزاد خواهد بود. بنابراین دستگاه بینهایت جواب دارد که جواب‌های آن به فرم زیر هستند.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -۲z + \frac{۱}{۳} \\ z - \frac{۱}{۳} \\ z \end{bmatrix}$$

سوالات عملی (۳۲ نمره)

تاریخ تحویل: ۸ اسفند ۱۴۰۱

پرسش ۱ (۳۲ نمره) همان‌طور که میدانید، کامپیوتر موجود بسیار نفهمی است و تمام تلاش خود را برای نفهم بودن می‌کند. یکی از مصادیق نفهم بودن کامپیوتر در نفهمیدن متون و شباهت میان آنها است. این است که برای فهماندن متن به کامپیوتر، چاره‌ای به جز تولید یک بازنمایی برداری از آن نداریم. یک بازنمایی برداری معروف در کامپیوتر از متون، کد ASCII و یا UTF برای ذخیره‌سازی کاراکترها و استفاده از دنباله‌ای از آنها برای ذخیره‌سازی یک متن است. اما مشکلی که این بازنمایی دارد این است که هیچ ارتباطی به بعد معنایی کلمات و جمله نخواهد داشت.

می‌خواهیم یک بازنمایی جدید برای جملات معرفی کنیم. فرض کنید که مجموعه‌ای از m جمله داریم که در مجموع از n کلمه متمایز تشکیل شده‌اند. حال از روی کلمات و جملات می‌خواهیم دو ماتریس زیر را تعریف کنیم.

(آ) ماتریس بسامد کلمات: هر چه یک کلمه در یک جمله پرتکرار تر باشد، احتمالا نقش پررنگ‌تری در معنی آن جمله خواهد داشت. ماتریس بسامد کلمات که آن را F می‌نامیم یک ماتریس $m \times n$ است که درایه i, j آن برابر بسامد کلمه j در جمله i است. به عنوان مثال در جمله

I like eating desert in desert.

۶ کلمه داریم و کلمه desert در آن دو بار تکرار شده است. بنابرین بسامد این کلمه در این جمله برابر $\frac{2}{6}$ است.

(ب) ماتریس متداول بودن کلمات: هر چه یک کلمه در جملات بیشتری واقع شده باشد، احتمالا بار معنایی کمتری دارد. به عنوان مثال حروف اضافی و ربط در اکثر قریب به اتفاق جملات واقع شده‌اند. ماتریس (بردار) تدوال کلمات یک ماتریس $n \times 1$ است که درایه iام آن برابر لگاریتم نسبت تعداد کل جملات به تعداد جملاتی است که شامل کلمه iام باشند. به صورت رسمی اگر D مجموعه کل جملات باشد و D_i زیر مجموعه از جملات باشد که شامل کلمه iام باشند، درایه iام ماتریس تداول کلمات که با I آن را نمایش می‌دهیم، برابر است با

$$I[i] = \ln \frac{|D|}{|D_i|}$$

حال بازنمایی جمله iام را به صورت ضرب درایه به درایه ردیف iام ماتریس بسامد در بردار تداول در نظر می‌گیریم.

در این مساله یک مجموعه به نام S داریم که شامل تعدادی جمله است. در مرحله اول باید بازنمایی این جملات محاسبه شود. سپس یک جمله query داریم که باید بازنمایی آن مطابق بازنمایی محاسبه شده برای جملات بالا محاسبه شود. در نهایت باید اندیس شبیه‌ترین جملهٔ S به query برگردانده شود. این شباهت از طریق محاسبه زاویه بدست می‌آید.

پیش از انجام هر یک از این موارد، باید علائم نگارشی متن را حذف کنید. همچنین باید از کوچک یا بزرگ بودن حروف صرف نظر شود.

ورودی

در خط اول ورودی عدد صحیح n وارد می‌شود که برابر $|S|$ است. در n خط بعدی در هر خط یک جمله از مجوعه S وارد می‌شود. در نهایت در خط بعدی جمله query وارد می‌شود.

خروجی

در خروجی باید اندیس جمله‌ای که بیشترین شباهت را به query دارد چاپ شود.

ورودی نمونه

```
3
This is the first document.
This document is the second document.
And that is the third one.
Is this the first document?
```

خروجی نمونه

```
1
```

پاسخ

```python
import numpy as np


class TF_IDF:
    def __init__(self, corpus):
        self.corpus = corpus
        self.corpus_size = len(corpus)
        self.corpus_dict = self.get_corpus_dict()
        self.corpus_tf = self.get_corpus_tf()
        self.corpus_idf = self.get_corpus_idf()
        self.corpus_tfidf = self.get_corpus_tfidf()

    def get_corpus_dict(self):
        corpus_dict = set()
        for text in self.corpus:
            for word in text.split():
                corpus_dict.add(word)
```

```python
        return list(corpus_dict)

    def get_corpus_tf(self):
        corpus_tf = []
        for text in self.corpus:
            text_tf = []
            for word in self.corpus_dict:
                text_tf.append(text.split().count(word))
            corpus_tf.append(text_tf)
        return np.array(corpus_tf)

    def get_corpus_idf(self):
        corpus_idf = []
        for word in self.corpus_dict:
            count = 0
            for text in self.corpus:
                if word in text.split():
                    count += 1
            corpus_idf.append(np.log(self.corpus_size / count))
        return np.array(corpus_idf)

    def get_corpus_tfidf(self):
        return self.corpus_tf * self.corpus_idf

    def get_query_tfidf(self, query):
        query_tf = []
        for word in self.corpus_dict:
            query_tf.append(query.split().count(word))
        return np.array(query_tf) * self.corpus_idf

    def get_similarity(self, query):
        query_tfidf = self.get_query_tfidf(query)
        return np.dot(self.corpus_tfidf, query_tfidf) / (
            np.linalg.norm(self.corpus_tfidf) * np.linalg.norm(query_tfidf)
        )


def get_input_sentence():
    return input().lower().replace(".", "").replace(",", "").replace("?", "")


def main():
    n = int(input())
    S = [get_input_sentence() for _ in range(n)]
    query = get_input_sentence()
    tf_idf = TF_IDF(S)
    similarity = tf_idf.get_similarity(query)
    print(np.argmax(similarity) + 1)


if __name__ == "__main__":
    main()
```